A Novel Focal Plane Wavefront Sensor with A Pure Photonic Approach

Mrs. Jin Wei

This thesis is presented as part of the requirements for the conferral of the degree:

Master of Philosophy (MPhil)

Supervisors:
A/Prof. Sergio Leon-Saval & Dr. Barnaby Norris & Dr. Christopher Betters

The University of Sydney
School of Physics

2023

Declaration

I, *Mrs. Jin Wei*, declare that this thesis is submitted in partial fulfilment of the requirements for the conferral of the degree *Master of Philosophy (MPhil)*, from the University of Sydney, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Mrs. Jin Wei

October 19, 2023

Abstract

Adaptive Optics(AO) is the major technique for observing celestial objects from ground-based telescopes and wavefront sensors (WFS) are one of the core tools for sensing the phase information of light. As Earth atmosphere distorts the wavefront (WF) of an incoming light from a distant point source, it is critical to know the degree of distortion of the phase in order to reconstruct the image at the diffraction limit for analysis. However, there are challenges for AO systems to resolve small and dim objects such as Earth-like planets next to their parent stars, mainly due to their small angular separation and high background glare from the stars. Current WFSs embedded in AO systems carry inherited problems that either are not able to correct particular types of wavefront errors (WFEs), or introduce more into the system due to non-common optical paths. This results in the inability of observing exoplanets and similar objects. It has been long realised that a new generation of WFSs is required to observe Earth-like exoplanets and similarly dim objects.

This thesis proposes and introduces a new type of all-photonic focal plane WFS. I will present the findings of using a 19 core photonic lantern (PL) as a novel type of focal plane wavefront sensor (FP WFS) in a laboratory setting. The aim of the experiment is to explore the feasibility of PL as FP WFS using a neural network (NN) in a physical setting. The PL, in theory, is able to function as WFS due to its ability to convert multi-mode (MM) inputs to single-modes (SM) outputs and vice versa. While direct detection of the input phase and intensity with a single PL (at a single wavelength) is not possible due to degeneracy, a NN is able to find correlations using non-linear function between the PL's inputs and outputs, hence establishes accurate predictions of the WF inputs. Chapter 1 provides information on the current adaptive optic (AO) technology used in Astronomy, and the presenting challenges of AO in ground-based telescopes. Chapter 2 introduces the two building blocks of the proposed technology, and the reasons why the combination of both PL and NN can restore the distorted WF. Firstly, PL as a device of interest with detailed explanation of propagation of light modes within the PL, i.e. how PL works as a WFS. Secondly, Chapter 2 also describes the neural network (NN) technique, its development, and its ability to solve problems autonomously. The actual experiment is detailed in Chapter 3 and 4. The results and analysis of the data obtained show a promising technique that is able to resolve WFEs that are currently extremely difficult for current WFSs. This demonstrates that PL as FP WFS are worth exploring further, especially in a real on-sky setting, and it can largely advance AO technology in general, from astronomy to free-space communications. Chapter 5 concludes the finding and future research direction.

Acknowledgements

Throughout this research project, it is impossible to complete it without the help and support from various people I encounter. I have learnt tremendous research knowledge through the teaching of my supervisors, A/Prof. Sergio Leon-Saval, Dr Barnaby Norris and Dr Christopher Betters. I have learnt much knowledge in academic research from this project and developed a strong liking in venturing further into this direction. I also gained highly valuable skills in coding and experimental procedures that I was unable to learn from my undergraduate years. Apart from all these, most importantly, I have gained insights that expand my horizon as a person on individual level. I have also received suggestions and exchanged ideas with my peers, Alison Wong, David Sweeney, Lucinda Lilley, Matthew Rahme and especially Louis Desdoigts who shared a code needed for analysing a part of the experimental results. I also would like to thank my partner, Liam Willoughby, who supported me throughout the project from the high and low. I also would like to thank my daughter, Meldoy Willoughby-Wei, who was born during my master project. You bring me so much more joy in my research and life.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AI	Artificial intelligence
ANN	Artificial neural network
AO	Adaptive optics
DL	Deep learning
DM	Deformable mirror
EM	Electromagnetic
FP	Focal plane
LWE	Low wind effect
MCF	Multi-core fibre
ML	Machine learning
MM	Multi-mode
MMF	Multi-mode fibre
NN	Neural networks
PL	Photonic Lantern
PSF	Point spread function
PP	Pupile plane
RAM	Random access memory
RMSE	Root-mean-squared error
RMS WFE	Root-mean-squared wavefront error
SL	Supervised learning
SM	Single mode
SMF	Single mode fibre
SNR	Signal-to-noise ratio
SR	Strehl ratio
TLU	Threshhold logcial unit
UL	Unsupervised learning
ULI	Ultrafast laser inscription
UTM	Universal Turing machine
VLT	Very large telescope
WG	Waveguide
WFS	Wavefront sensor

Chapter 1

Introduction

Astronomy, being an observational science, involves the study of celestial bodies and the analysis of their motion and composition to infer their evolution over time. When it comes to observing distant objects, light in the form of electromagnetic(EM) waves exhibits a planar wavefront with uniform intensity, assuming the source is located at a considerable distance [1]. However, ground-based observations face a significant challenge due to atmospheric turbulence, which causes distortion in the WF. In astronomy, the concept of *Seeing* pertains to the angular resolution limited by atmospheric conditions [2]. One potential solution to mitigate this issue is to deploy telescopes in near-Earth orbits, as exemplified by the James Webb Telescope and the Hubble Telescope. However, this approach is accompanied by exorbitant costs for setup and maintenance. Additionally, space telescopes are constrained by factors such as their smaller size compared to their ground-based counterparts, which limits their overall functionality. Another effective approach to mitigate atmospheric turbulence is the utilization of AO. AO is a system that integrates various optical and mechanical components to correct distortions in the optical wavefront through phase conjugation methods [2]. The concept of AO was initially proposed in 1953 and further developed during the latter part of the 20th century [3, 4]. With the continuous advancement of AO technology, ground-based telescopes have been able to achieve results comparable to their space-based counterparts [3]. Furthermore, AO plays a crucial role in modern telecommunications [5], including satellite communication between near-Earth orbit and the Earth.

AO systems employ a synergistic combination of hardware and software components to measure distortions and promptly apply corrective measures in real-time [2–4, 6]. WFSs are pivotal hardware components within AO systems and have undergone significant development alongside the advancement of AO technology over the past six decades [2]. These sensors play a crucial role in measuring and characterising the distortions present in the incoming wavefront. By accurately capturing the aberrations, WFSs enable precise and real-time corrections to be made. As a testament to their effectiveness, WFSs have been successfully integrated into numerous state-of-the-art telescopes worldwide. This integration has greatly enhanced the observational capabilities of these telescopes, allowing astronomers to capture clearer and more detailed images of celestial objects.

This thesis is focused on researching into a new type of WFS using a pure pho-

tonic approach. Such an approach to WFS has the potential to provide a number of benefits over traditional optics setups that use bulk components. One of the primary advantages is the ability to downscale the size of the equipment required, making it easier to integrate into other systems and/or to make it portable for field applications. Additionally, photonic approaches can reduce the complexity and cost of WFSs, since they rely on fewer components and can be fabricated using standardised, mass-production techniques as the fabrication of photonics advances.

Overall, a pure photonic approach to WFSs has the potential to revolutionise the field of wavefront sensing, offering new opportunities for research and development in a range of areas, from astronomy and remote sensing to manufacturing and biomedical imaging.

1.1 The Turbulent Atmosphere

In order to understand AO technology, it is helpful to understand atmospheric turbulence and its effect on an incoming wavefront. Atmospheric turbulence is generated by mechanical processes and the thermal instability of the atmosphere [3]. The day-night heating and cooling cycle causes convection currents and winds. These currents and winds flow over uneven topography, such as forests and mountains, leading to regions of air with different refractive indexes [3]. The wavefront of light is distorted when travelling through these air variations. The Kolmogorov-Obukhov scaling law dictates that the proportionality constant, C_n^2, also called the index of refraction structure constant, describes the fluctuations of the refractive index in atmospheric turbulence. It characterises the strength of turbulence and is a key parameter determining seeing quality and the performance of AO systems [3]. A larger C_n^2 indicates a more turbulent atmosphere [2, 3].

An AO system faces three environmental constraints from the Earth's atmosphere, Fried's coherence length, isoplanatic angle, and the Greenwood frequency [2, 3]. Fried's coherence length, r_0, which is the parameter that describes the spatial extent of a section of turbulence [3] as a function of wavelength, where

$$r_0 \propto \lambda^{\frac{6}{5}}. \tag{1.1}$$

For visible light, r_0 is equivalent to a few centimetres in magnitude [2]. The larger the value of r_0, the better the observation condition is [2], as it inversely relates to C_n^2. This is is described by,

$$r_0 = [0.423k^2 \int (C_n^2(z), dz)]^{-\frac{3}{5}}, \tag{1.2}$$

where k is the wavenumber of the free space as $k = \frac{2\pi}{\lambda}$, and C_n^2 is related to the altitude z (in kilometres). It is such an important parameter in astronomy that the term *Seeing* is defined by the ratio of the incoming wavelength, λ over r_0. A value of r_0 less than 5 cm at the observation site is said to have a poor seeing condition, while a value greater than 20 cm is said to have a good seeing condition [6].

The isoplanatic angle, θ_0, is the maximum angle allowed for a measurement of the wavefront between two points at the same altitude with less than one radian root

mean square (RMS) error in phase from each other [5] and is given by,

$$\theta_0 = 0.31(\frac{r_0}{L}). \tag{1.3}$$

As seen in Eq. 1.3, θ_0 is proportional to r_0 over the propagation path, L, the distance between the altitude of the measured source and the telescope. As the atmosphere changes rapidly over time, the term Greenwood frequency f_G is used to describe the temporal element of the atmosphere, as seen in Eq. 1.6. It is a useful parameter that defines how well the AO system can process changing turbulence in the temporal domain. The Greenwood frequency, f_G , is highly dependent on the coherence time, $\tau_0 = 0.314\frac{r_0}{V}$ (where V is the wind speed assumed at constant), which characterises the temporal change of the atmosphere turbulence across a telescope [3].

f_G can be expressed as integration over layers of turbulence, z, with a constant wind speed, V, along the path of $\sec(\beta)$, where β is the zenith angle:

$$f_G = 2.31\lambda^{-\frac{6}{5}}[\sec(\beta)\int C_n^2(z)V^{\frac{5}{3}}(z)dz]^{\frac{3}{5}} \tag{1.4}$$

$$= 0.43(\frac{V}{r_0}). \tag{1.5}$$

It can also be written in relation to the coherence time, τ_0, as:

$$f_G = 0.134/\tau_0. \tag{1.6}$$

Therefore, these three factors largely constrain the condition of the environment in which an AO system operates. However, despite these well-established mathematical models, the AO system cannot compensate for all turbulence. This is due to limiting factors such as complex and rapid changes in the atmosphere, as well as the systematic error existing in the AO system itself, for example, the processing speed of the AO system and non-common-path (NCP) aberrations [2, 3]. Some scaling laws are developed and implemented to evaluate the performance of AO systems [2], which are detailed in Sections 1.2 and 1.3.

1.2 How Adaptive Optics works

There are three principal components in an AO system, a WFS, a deformable mirror (DM), and a control computer, which are integrated with a telescope and an optical system [2, 5], as seen in Fig. 1.1. Atmospheric turbulence causes optical path difference due to slight change in refractive indexes of the air [2, 6] resulting aberrations. These aberrations cause distortion, blurring, and other image quality issues and is called wavefront error (WFE) [2, 6–8]. During the on-sky operation, a bright background star or an artificial guide star, created by a powerful laser or Rayleigh beacon, is usually used as the reference star [2], so that the wavefront contained in the incoming atmospheric turbulence can be sampled and calculated in real time by a ground-based WFS.

The WFS detects and interprets the phase component in a beam's electric field since any light beam can be presented as a wave function with amplitude, A, and

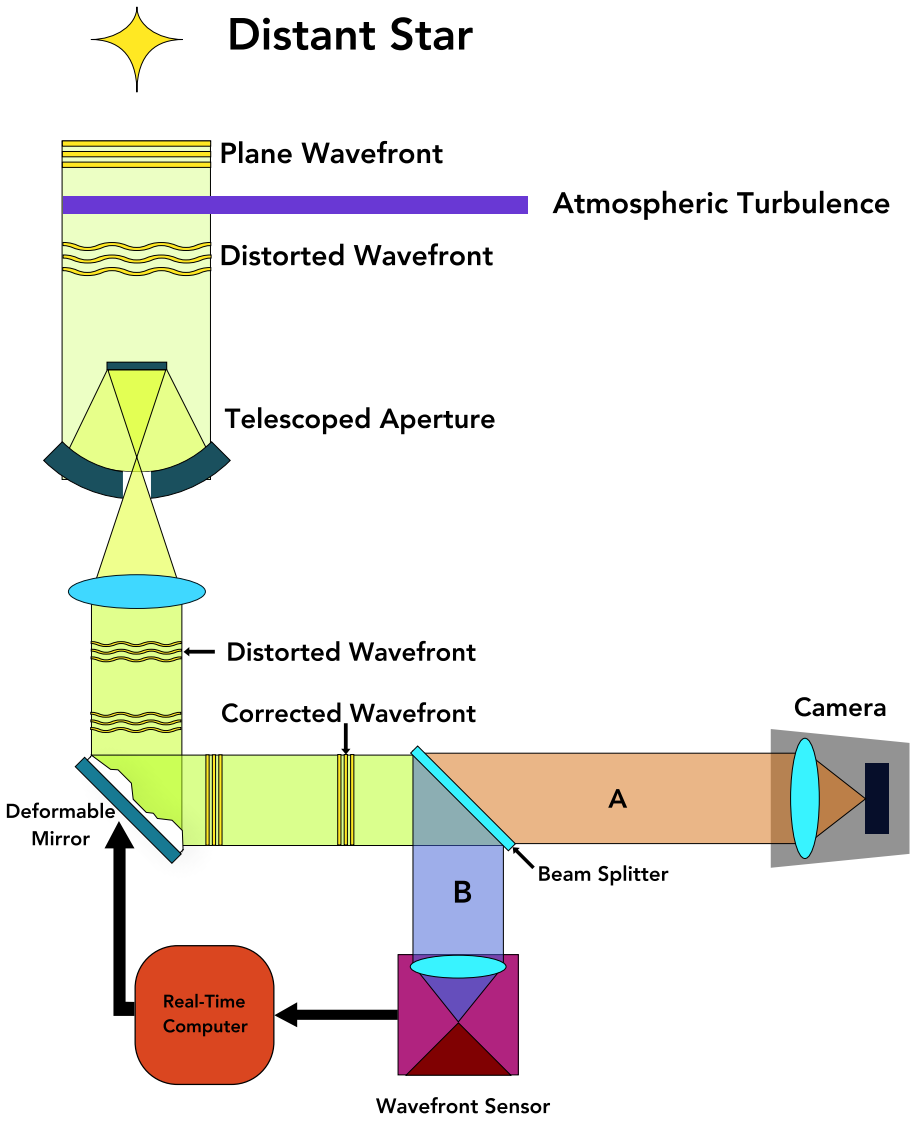


Figure 1.1: A schematic of a typical AO system with a pupil plane WFS: The planar wavefront is distorted by atmospheric turbulence. The DM corrects the distortion by phase conjugation in conjunction with a real-time computer system. The WFS plays a crucial role in detecting phase distortion and gives a quantifiable value to such distortion, hence, the computer can relay the information to the DM for correction. Noted that the optical train A and B are schematics, in reality, each contains many different optical components. Due to the position of the WFS at the object's pupil plane, a beam splitter is employed to produce two optical trains for the WFS and the scientific camera respectively.

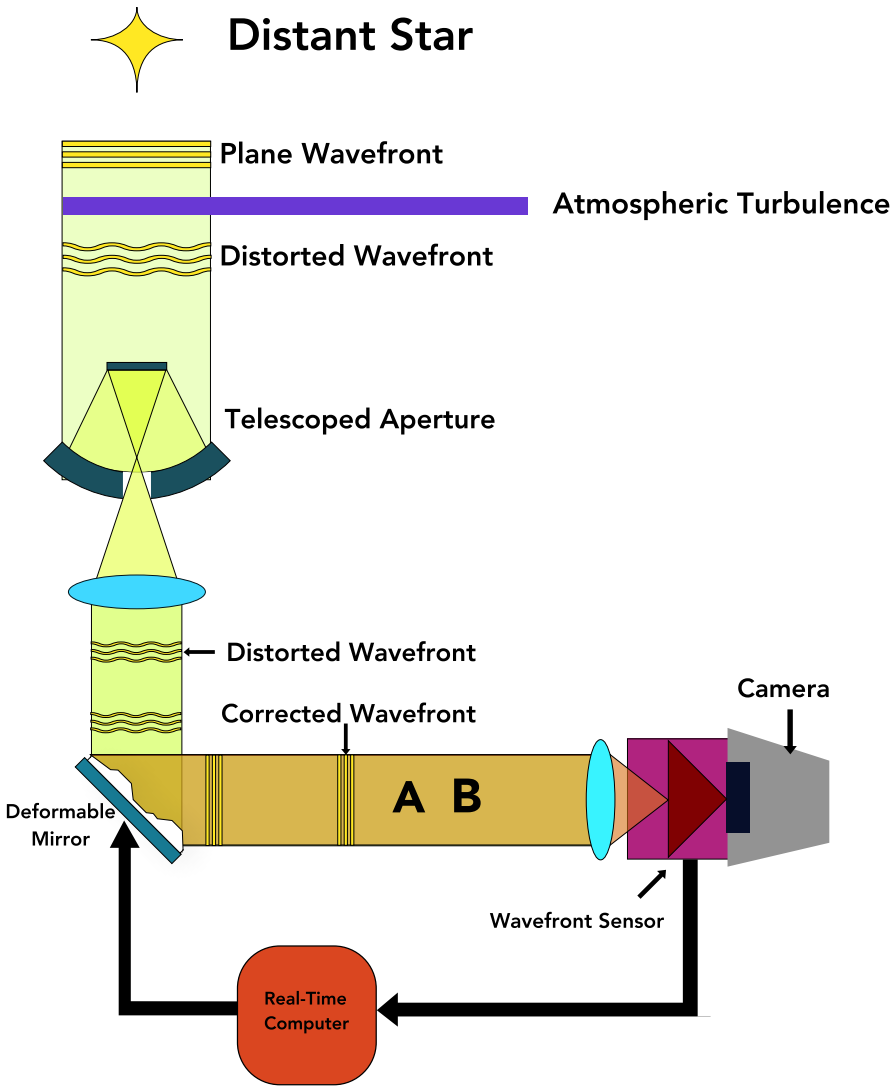


Figure 1.2: A similar schematic of an AO system but with a FP-WFS instead: Noted that the optical train A and B are now one of the same path due to the fact that FP-WFS, sensing at the focal plane, does not require a separated optical train for analysis unlike in the ones in Fig. 1.1. This design is currently not in existence.

phase, θ; the general function of the said electric field is: $U = Ae^{i\theta}$[2]. The control system, a computer, applies the complex conjugate of the function, $Ae^{-i\theta}$, onto the DM. The two terms, $Ae^{i\theta}$ and $Ae^{-i\theta}$, mathematically cancel out to give the function U, with a phase of 0. This method is called phase conjugation. In an ideal situation, AO is able to restore the distorted image to its ideal diffraction limited (absent of phase distortion) point spread function (PSF), which is the mathematical function that describes how a perfect point source of light, such as a star or a point on a printed image, is spread out or blurred by an imaging system [9]. However, there are a few limiting factors that affect the performance of an AO system.

The term Strehl ratio (SR), SR, is introduced to describe the imaging performance of an AO system [2, 3, 5, 10]. SR is defined in terms of the ratio of the maximum intensity in an aberrated point source image to the maximum intensity of the image at the diffraction limit [4, 10]. SR describes how well the optic system can image the object to the said object's diffraction limits. SR is thus indirectly established from Fraunhofer integral [2], as the later is a mathematical formula that describes how a complex wavefront propagates from an object to a distant observation plane in the far-field regime. The French physicist, Augustin-Jean Fresnel was the first to develop this function in Fraunhofer diffraction theory [11]. The peak intensity I_0, of an incoming wave with wavelength, λ, and total power, P, through a circular aperture with a diameter, D, without any phase aberrations (i.e. diffraction-limited), is

$$I_0 = P\frac{\pi D^2}{4\lambda^2 R^2}, \tag{1.7}$$

where R is the propagation distance. This diffraction-limited image is perceived physically as a series of concentric rings surrounding the central point of light with 80% of the total intensity, called an Airy ring. Phase aberrations are present when there is a mismatch between the actual wavefront and the theoretically perfect wavefront, also known as WFE. As a result, the maximum intensity, I_0, decreases in intensity. Given that the wavefront error, σ, is known, the actual intensity, I, can be derived from 1.7 as

$$I = I_0[1 - (\frac{2\pi}{\lambda})^2\sigma^2]. \tag{1.8}$$

Noting that the part in squared bracket in Eq.1.8, is the SR component [2] which has a value between zero and unity. If the SR component is at unity, it represents an optical system free of aberration [10]. SR can also be written in exponential expansion terms [2] as

$$SR = \exp[-(\frac{2\pi\sigma}{\lambda})^2] \tag{1.9}$$

. A more in-depth explanation of SR and image quality can be found in [10]. If $SR \geq 80\%$, the AO system is said to be at the diffraction limit [6]. From Eq. 1.9, a signal with a shorter wavelength is ideal, such as visible spectrum, as it tolerates a greater WFE to achieve the ideal 80% in SR. However, in practice, the signal coming from the distant star usually consists of a longer wavelength as the Earth atmosphere tends to commonly scatter the shorter wavelength of light more (as opposed to the longer wavelength) [12] . As a result, most ground-based telescopes utilise longer wavelength, such as near-infrared, as the operating basis. This constrains the range of the residual WFE that is allowed for a good AO system. Notwithstanding, in

recent years, there has been an increased interest in developing an AO system for shorter wavelength (700 nm and less) [5, 13]. The shorter wavelength provides higher angular resolution as these two terms are proportional to each other. A higher angular resolution provides opportunities for observation of stellar physics, close binary stars and Earth-like exoplanets etc. [13].

1.3 AO performance

Challenges for an AO system during real-time operation are also present, since no current AO system can correct all the little turbulence in the air [2] . The performance of AO systems is evaluated by several different metrics [5] and for the focus of the thesis, error arising from WFSs are discussed below.

Ref.-[14] discussed AO errors in their 2004 paper and identified that WFS noise halo is the only fundamental limitation for extreme AO (ExAO), an AO system that is specialised in optimising wavefront quality over a narrow field centred around a bright star [13]. Ref.-[14] also proposed resolutions to address the errors mentioned above, such as a high density of actuators ($d \approx 10 - 20cm$) and a fast AO system ($f \approx 1 - 2kHz$) to reduce WFS noise; a low-aliasing WFS to resolve aliasing error; and adding an analytical process to remove cophasing errors.

In astronomy, the most widely used metric for evaluating the AO system is the delivered SR, or root-mean-squred(RMS) wavefront [3]. The delivered SR relates to the residual wavefront error variance, σ^2_{WFE} by Marechal approximation [5]:

$$S \sim e^{-\sigma^2_{WFE}} \tag{1.10}$$

The σ^2_{WFE} is the sum of a few uncorrelated contributors [5]:

$$\sigma^2_{WFE} = \sigma^2_{fitting} + \sigma^2_{reconstruction} + \sigma^2_{bandwith} + Other \tag{1.11}$$

The three main contributors in Eq. 1.11 are: the fitting error,$\sigma^2_{fitting}$. The reconstruction error, $\sigma^2_{reconstruction}$, which is the sum of all the effects that reduces the wavefront measurement accuracy, such as measurement noise, calibration noise, sampling errors, aliasing and chromaticity [5].

However, these errors can be complicated to fix, as solving one error may often exasperate a different error[3], AO systems are often optimised in selected features. They are designed according to their required functionalities.

The fitting error, denoted as $\sigma_{fitting}$, can be attributed to two sources: the DM and the WFS [3, 6, 15]. In a DM, numerous small mechanised units, called actuators, are controlled by a closed-loop algorithm to manipulate the surface shape of the DM. The resolving power of the DM in fitting spatial structures is determined by the total number of actuators it possesses [2, 15]. However, when the wavefront distortion becomes highly complex, the DM may encounter difficulties in accurately fitting the wavefront due to its limited number of actuators.

WFSs that employ linear approximation as the basis for fitting spatial distortions are powerful tools. However, it is important to note that linear approximation may yield sub-optimal results when attempting to fit highly non-linear modes of the complex wavefront.

Consequently, the fitting error arises from the insufficiency of the DM and/or WFS to accurately fit the spatial structures of the incoming phase, primarily due to the complexity of its distortion. A third source of the fitting error is the WFS's inability to fully resolve certain wavefront aberrations. Many currently deployed WFSs lack the capability to detect or are insensitive to specific wavefront aberrations. For example, the low wind effect (LWE), which is a type of aberration that existing WFSs are unable to detect [16–18]. Another example is non-common-path (NCP) aberrations, arising from differences in the optical path between the scientific channels and the WFSs, resulting in the failure of the WFSs to detect these aberrations [3, 7, 19].

1.4 Wavefront Sensors

A WFS is a device used to measure the aberrations in an optical wavefront [2, 3, 5, 6, 13, 15]. The performance of the AO depends on the accuracy of the WFS that detects the incoming wavefront, in conjunction with various reconstruction algorithms[20].

A WFS typically works by comparing the incoming wavefront to a reference wavefront or a known ideal wavefront. The sensor detects the phase shifts or deformations in the wavefront and provides feedback to the AO system, which applies the necessary corrections to the optical system to compensate for the aberrations.

1.4.1 Pupil Plane WFS

Pupil plane wavefront sensing (PP-WFS) is a common technique used in AO systems to measure the wavefront aberrations. The technique involves splitting the incoming wavefront using a beam splitter and projecting the resulting two beams respectively onto scientific channels and a WFS, seen in Fig.1.1, such as a Shack-Hartmann [21], a pyramid sensor[22] or the curvature WFS [4], which measures the phase shifts in the wavefront. The beam splitter is typically placed in the pupil plane of the optical system, seen in Fig. 1.1, which is the plane where the incoming wavefront passes through the aperture of the telescope or other optical instrument, hence, the name given to this type of WFS as PP-WFS conventionally. There are several disadvantages when using a PP-WFS in AO systems. Firstly, PP-WFSs can not resolve the NCP aberration. NCP aberrations are the result of non-common optical components traversed by the two split beam paths by a dichromatic beam splitter (seen in Fig. 1.1), which causes the differences in optic path between the wavefront seen by the WFS and the one used to observe the image [7, 8, 19]. Conventional AO systems incorporating PP-WFSs unavoidably give rise to NCP error because of the inherited alignment of the optical components. In Fig. 1.1, beam path A and B does not share a commonality and this difference in the two paths introduces the NCP aberration into the system. Aberration can take low or high-order forms and results in speckles appearing in the final images [23]. These speckles, particularly formed from low-order forms, are harmful when a coronagraph is used to form the final images [20] as they distort the image more and create false features. Speckle noises are often random and difficult to correct [13]. These aberrations are not seen

by the PP-WFS, and thus, cannot be corrected. This is one of the major limiting factors in the high-contrast Extreme AO systems [8, 17]. NCP aberration errors also affect the proper calibration of a WFS, which is an important step for closed loop performance of the AO system [6].

In Ref.-[19], a focal plane sharpening (FPS) method is applied in conjunction with phase diversity (PD) to correct the NCP aberration. However, this method requires a large number of images and time to achieve the desired outcome. The applied NCP aberration must be small (-0.7 to 0.7 radians) in incident RMS WFE with the final reconstruction root-mean-squared error (RMSE) of 0.4 in radians. The PD method described in [19] also has its disadvantages, such as its limitation on the types of aberration it can resolve. Furthermore, the characteristics of the aberration must be accurately known before correcting those same aberrations. PD also requires highly optimised algorithms, which demands a large portion of computation time; this greatly limits the practical application of FPS PD in correcting NCP aberrations by increasing the complexity in both software and hardware requirement.

Secondly, PP-WFSs are insensitive to aberrations called the Low Wind Effect (LWE) [16–18, 24]. LWE, also named island effect (IE), is a type of turbulence that occurs at the telescope top end. It is attributed to radiative cooling of the telescope spider (the structure holding the secondary mirror, seen in Fig. 4.17) caused by temperature differentials in its surroundings [16, 18]. As the temperature differential has an effect on the refractive index of the air, the wavefront of the incoming light becomes distorted due to the inhomogeneous refractive indexes of the layers of air within the telescope light path [16–18, 25]. Due to the fact that the effect is largely dominating when the wind speed can not blow away the thermally effected air pockets during low wind conditions, the heat transference from the spider to the surrounding air is more efficient at the lower wind speed, it is so named the " low wind effect" [16]. These aberrations produce a strong effect in the image plane, see the bottom left image in Fig. 4.17. LWE aberrations can also result in speckle noises [13] , which represent a significant challenge in high-contrast extreme adaptive optics (EXAO) systems used for exoplanet hunting. Speckles occur randomly and can resemble the angular size of exoplanets, making it challenging to distinguish between them [13]. Furthermore, the pure randomness of speckle occurrence, combined with their similarity in angular size to exoplanets, increases the likelihood of mistaking a speckle for an actual exoplanet. This can introduce false positives or false detection in astronomical observations, potentially impacting the accuracy and reliability of exoplanet detection and characterisation. The presence of LWE aberration has been reported in Spectro-Polarimetric High-Contrast Exoplanet Research (SPHERE) [16]. The current WFSs mounted on the telescopes are insensitive to the shearing in phase (seen in the bottom right of Fig. 4.17) that result in these aberrations. As a result, there is active AO research in relation to new types of WFS that are capable of detecting LWE.

There are also other issues for PP-WFSs, such as poor sensitivity at low spatial frequencies and the requirement of a large number of photons to work [26], which is detrimental to exoplanet imaging. A study by Ref.-[26] has shown that the curvature WFS, a form of PP-WFS, displays promising results in ExAO simulations by enhancing sensitivity. However, one limitation of this WFS type is its inability

to address the phase wrapping [26], a jump or discontinuation of phase when incident phase exceeds 2π causing ambiguity in wavefront measurement. Furthermore, curvature WFS still suffers from NCP aberrations in the study of Ref.-[26].

1.4.2 Focal Plane WFS

Focal-plane wavefront sensors (FP-WFSs) measure the wavefront in the focal plane of the objective, where the two-dimensional Fourier transform of the object is contained [27]. FP-WFSs can eliminate the NCP issue and is highly applicable to phase diversity methods by directly using Zernike polynomials to fit individual aberrations at focal plane [28]. FP-WFSs eliminate NPC errors by sampling wavefront at the same plane as the detectors [13], rather than the need to use additional components to redirect light, such as a beam splitter. FP-WFSs are sensitive to aberrations caused by the low-wind effect (LWE) as these aberrations have a strong effect in the image plane. The aberration caused by LWE can not be detected by a PP-WFS, such as Shack-Hartmann wavefront sensor, and a phase-measuring sensor at the focal plane may be needed to detect such aberrations. This is because PP-WFS measures the gradient of the wavefront, and does not directly measure the phase of the wavefront [2, 16, 17, 20]. Section 4.3 will further discuss on this aspect.

In Ref.-[29], they also discussed the advantages of a FP-WFS over its counterpart, a PP-WFS, with higher accuracy and dynamic range Peak-to-Valley (PV). Furthermore, they compared FP-WFS with interferometers and concluded that FP-WFS was superior in overcoming the effects of vibrations and environmental disturbances. Therefore, FP-WFSs are desirable in future AO systems.

However, the detection of a wavefront in FP can be complicated due to limitations of current imaging detectors, such as CCD or CMOS chips, which only detect the intensity I, of the beam, instead of the phase [20]. This is due to the complex wave function of light described in Eq. 1.12 (ω is the angular frequency, t is time, β is the propagation constant of the wave and z is the orthogonal axis of the x axis) and the intensity function, I, is the modular function of Eq. 1.12 as 1.13:

$$E_{(x,z,t)} = \Psi(x)e^{i(\omega t-\beta z)} \tag{1.12}$$

$$I = \mid E_{(x,z,t)} \mid^2 \tag{1.13}$$

Therefore, the phase information,$i(\omega t - \beta z)$, of the signal is lost for conventional imaging devices such as a camera.

There have been many attempts to achieve FP wavefront sensing in order to bypass this problem, such as phase diversity methods [30], Fast and Furious method [31], and the Zernike Asymmetric Pupil Wavefront Sensor[8]. However, most of the methods rely on a linear approximation or a small wavefront error ($\ll 1$ radian)[20]. Furthermore, these FP-WFSs suffer greatly from the chromaticity problem. It should be noted that these FP wavefront sensing methods have a significant drawback, as highlighted in Ref.-[20]. Specifically, these methods are unable to seamlessly incorporate spectrography into their functionality, which is a critical aspect of advanced exoplanet characterisation.

In this thesis, an all-photonic device, the Photonic Lantern (PL), is tested as a new type of FP-WFS, which is placed at the image plane of the objective lens to

directly measure the amplitude and phase of the image without the need for linear approximations in a laboratory setting. In Chapter 2, we look at the theory and properties of an PL and how the PL is manufactured. In Chapter 2, we explain the function of a neural network (NN) which is employed for the reconstruction of the wavefront measured by PL using the non-linear method. In Chapter 3, we describe and analyse the experiment that demonstrates the PL operating as FP-WFS in a laboratory setting, in conjunction with a NN to process data recorded to achieve the reconstruction of the input phases. The PL also can be augmented with a spectrograph (as it did in the experiments described in this thesis), which has great implications for advanced exoplanet analysis, as mentioned above, to combat issues such as speckles and phase wrapping in modern AO technology.

Chapter 2

Photonic Wavefront Sensors Powered by Neural Networks

In this chapter, there are two key components for building the WFS proposed and demonstrated in this thesis: the photonic lantern (PL) and Neural Network (NN). The PL - its multi-mode (MM) to single-mode (MM) conversion properties- is used as the light sensing mechanism to interrogate and predict the incoming WF. However, due to the highly non-linear relationship between the WF, it is difficult to produce a function that can resolve the problem. NN is thus used to map the relationship between the intensity measured from the PL's output and the phase at the PL's input. The detailed information and history of the both are described in Section 2.1,2.2,2.3 and 2.4

2.1 Photonic Lantern

Photonics is the field that describes the application and usage of light, or photons, to transmit and/or process information, or modify materials [32]. Photonics includes optical sensors, optical imaging and processing, as well as optical communications, which is one of the most important technologies of the 21st century [32]. The hallmark of modern photonics is based on the quantum property of light which was explained by famous Einstein's photoelectric effect in 1905. Photonics is closely related to optics which is a much older stream of physics, and optical fibres developed during the late 70's are used in many photonic fields as sensors or light transport methods.

An optical fibre is a cylindrical dielectric waveguide (WG) made of low-loss materials, such as silica, with a core surrounded by cladding [33], the light is guided within the core due to total internal reflections [33]. The concept of optical fibres was discussed in Tyndall's lectures in the late 1900 century [34], and materialised this idea into the prototypes of optical fibre[34]. Due to the low-loss nature of optical fibres, it quickly replaced the copper coaxial cables in telecommunications [33].

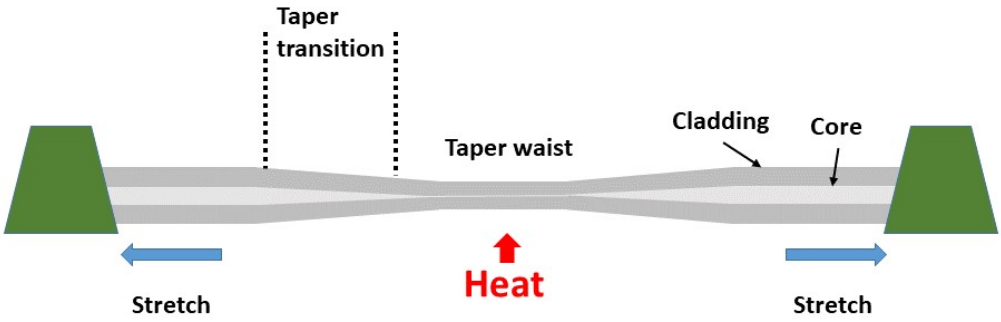
The introduction of optical fibres in astronomy as a means of transporting incoming light to various instruments for analysis started in the 1980 [20, 35–37] enabling the downscale of certain equipment in astronomy. For example, before incorporating optical fibres as means of light transportation, instruments such as spectrographs

had to be at the focus of the telescope, which limited the size of the spectrograph when integrated into the telescope system. With the popular usage of optical fibres in the astronomy community, so has the incorporation of photonic technologies, such as fibre Bragg gratings (FBG) and the fibre couplers, in the early 2000s [38]. The field of Astrophotonics started in early 2000's as a bridge between astronomy and photonic processes with the primary goal of harnessing technologies developed in photonics to enable a new generation of astronomical instrumentation [38].

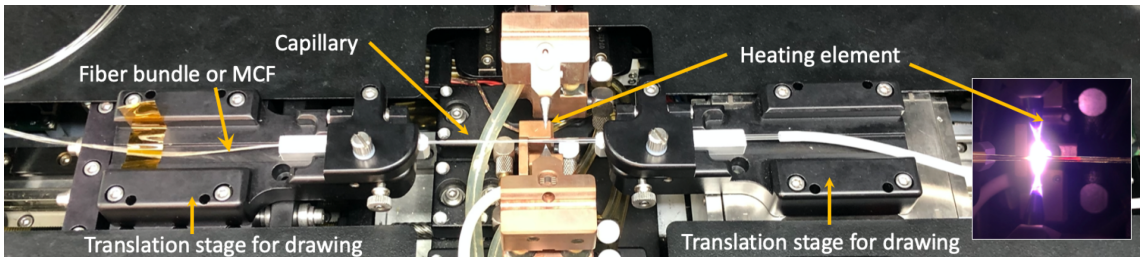
The first PL was made in 2005 [35–37] in search of a device that was capable of covert MM to SM light with low loss and benefited from the attributes of both multi-mode fibres (MMFs) and single-mode fibres (SMFs). Due to the wide use of optical fibres in astronomy, it was very well understood that MMFs with a large core size were able to capture more light from the stars, which was very important in the study of very faint objects, such as exoplanet hunting. However, SMFs were essential in applying new photonic functions, such as filtering the noise from science signals for analytical purposes [35–37]. Unfortunately, the desired modal guidance characteristics of SMFs and MMFs were not interchangeable, and they could not simply be connected to each other without information loss during the transition. Most photonic technologies that use SM fibres, such as FBGs, could not be directly applied to MMF, which were proposed earlier on for sky background suppression [36]. Hence, a solution for applying SM photonic functions to MMFs was indeed needed. There were some attempts to solve this problem, for example, Sun et al. performed a narrowed MMF grating [39], however, the MMF and modes used were highly specialised and were incompatible with the general MMFs. In 2015, PL were proposed and designed to achieve efficient conversion of MM to SM conversion and vice versa via a low loss optical transition device [36, 37, 40]. Consequently, PLs allowed for SM performance in MM fibres, such as more efficient data collection and processing using photonic technologies was possible.

Although the PL has been used in astronomical instruments as an MM to SM converter, its function as a FP-WFS has not been fully explored in a laboratory setting, indeed this thesis reports on the first real demonstration of such technology. Previously, using a digital model of a PL to test its function as a FP-WFS through computational simulation [41] was carried out in a paper published in 2016. Ref.-[41] also discussed several advantages of using PL as WFS, such as a PL could be tailored in any shape to match the configuration of expensive sensors so that it reduced the wastage of unused pixels area on expensive sensors due to traditional shapes of WFS issues. However, those attempts did not use them as proper WFSs but more like spatially non-degenerate fibre splitters. Thus, not taking advantage of one of the most important features of PL - that it can decouple a MM source into its subset of SM with very low loss will maintain the phase information from the source during the PL coupling process. The composition of different guided modes at the MM end of the PL will transform in a set of SM outputs of the PL with different intensities distribution depending on the incident phase. Hence, this information can be used to infer the phase information of the source at the focal plane (PL input); in other words, the PL can be used as WFS at the focal plane (FP).

The advantages of using PLs as WFSs are: facilitating the incorporation of other photonic technologies directly into the AO system; It enhances the AO system's



(a) Schematic of a taper rig for tapering an optic fibre: As the fibre being stretched, the diameter at the centre part of the body became smaller. Once the fibre is tapered, it is cleaved into two: a PL and a sacrificial fibre (the end to be discarded).



(b) A top view of the taper rig machine, 3SAE.

Figure 2.1: Tapering rig for the production of PLs.

ability to minimising the interference of speckles, produced by monochromatic source [42], using broadband source as well as open the gate for multipurpose analysis using other photonics.

2.2 How the Photonic Lantern is made

Early on PLs were made by inserting SMFs into a glass preform cane with a 19-hole array, and the cane was drawn to MM photonic crystal fibre (PCF) [36]. Ref.-[36, 37] showed that this device was capable of producing a MMF system with SM photonic functions of a SMF - e.g. SM FBG response in a MM system. The downside of their work was the high loss between the MM-SM transition, due to the mismatch between the number of SM cores in the MCF end and MM in MMF end, and the lack of practicality of the manufacturing method in mass production [36, 43] due to the air-clad PCF used in MM. Leon-Saval & Birk proposed another production method to simplify the fabrication and achieve low loss for similar devices [36] by tapering down a multi-core fibre containing an array of SM cores. The method was used in the fabrication later on of PLs with larger number of modes and SM WGs.

Fibre tapering is the optical fibre post-processing method that makes PLs and

gives them all of the desirable characteristics. Fibre tapering was first used in 1981 by Ref.-[44] when a heat source was applied to a SM fibre while the fibre is stretched by some mechanical means under a control condition [45]. The process of fibre tapering was further explored in more mathematical detail by Birks and Li in 1992 [45], so a rig could be made to specialising in tapering fibres of desired features; see Fig. 2.1a fibre tapering has been used to produce couplers, such as the PL, which fundamentally is a fibre coupler, and many other applications, such as fibre sensors of optical and acoustic detection. Nowadays, PLs are made by commercially available glass post-processing equipment, e.g. Vytran, 3SAE, Nyfors, etc.

There are several ways of making a PL [46], and those methods are listed in Fig. 2.2. A standard PL is made with a bundle of SMFs placed inside a capillary that functions as the secondary cladding with a refractive index lower than that of the core and cladding of the SMFs[38]. The bundled fibres then undergo the tapering process using a specialised mechanical apparatus, see Fig. 2.1a to reduce the overall diameter of the cross section. The cores of the SMFs are reduced to such a size that the SMFs cores become ineffective WGs; the cladding of the SMFs is fused together into one structure. At the end of the taper transition, a new MM core is formed, with the secondary cladding becoming the cladding and the now fused SMFs being the core. The tapered region is then cleaved and the cleaved surfaces polished to produce a PL with one MM end and one SM end. As a result, PL is a monolithic device that allows both SMF and MMF capacities (seen in Fig. 2.6) and each SMF is at the diffraction limit [35].

PL produced with multicore fibre (MCF) is an easier way to produce, since they do not suffer from fabrication imperfections due to early fusing and fibre twists during the SMF bundling. A MCF, although the fibre itself requires a custom optical fibre draw, it can be obtained in kilometres of length. Furthermore, MCF PLs allow easier scalability in the number of modes for individual PLs. Thus this method makes the manufacturing effort easier.

However, the downside of PL production is that each PL is produced by hand, which is time-consuming. The ultra-fast laser inscription (ULI) technique has been tested as a production method for PL [38] which uses laser writing to create PL in a single piece of glass block, see details in Fig. 2.3. The adiabatic transition is achieved by gradually bringing the SM WG together so that they are strongly coupled with each other. This is much more practical in mass production with the additional advantage of size reduction for PL. The disadvantage of such a method is the difficulty of integration with standard optic systems available in the current market [38].

2.3 Principle of the Photonic Lantern

The principle of PL is based on a very gradual change, also known as an adiabatic transition, in the tapering region of the PL, compared to the incoming wavelength; The MM in the MMF end decouples into its subsets of SMs at the MCF end consisting of an array of identical SM cores, as the modes propagate along the body of a PL. If the number of degrees of freedom, i.e. the number of modes, between the MM end and the SM end of the PL, hence a low loss optical transition will be

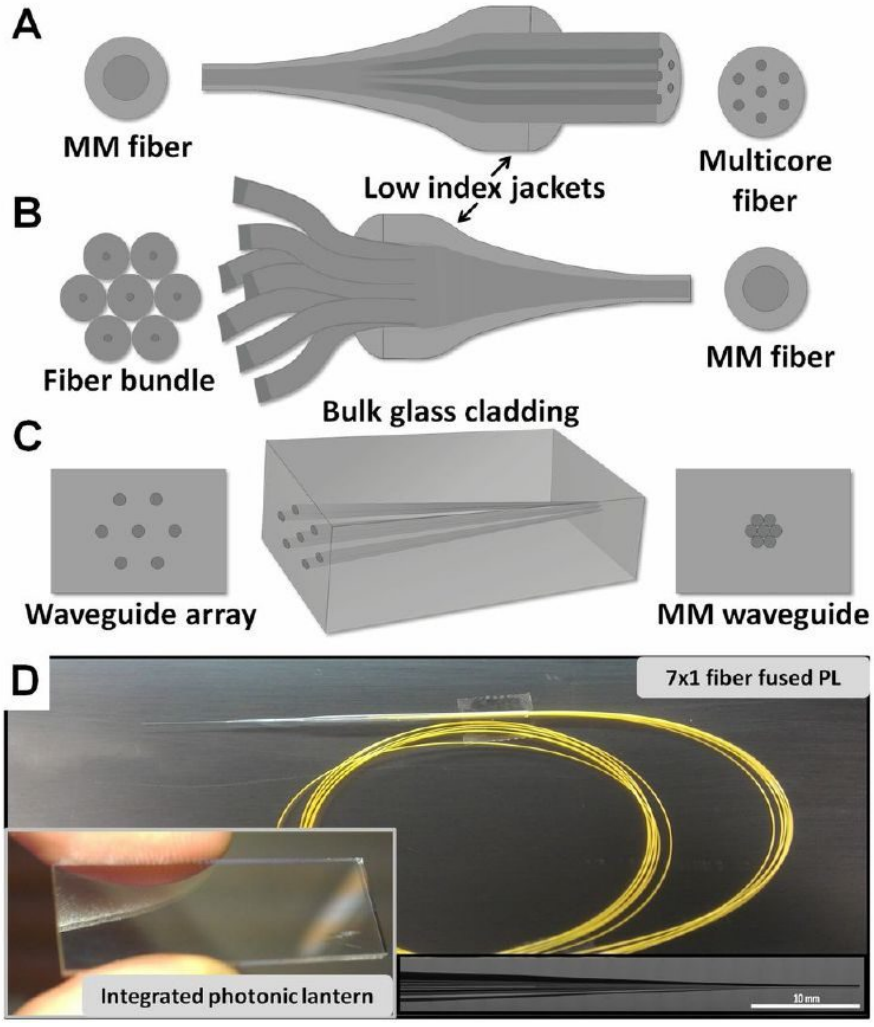
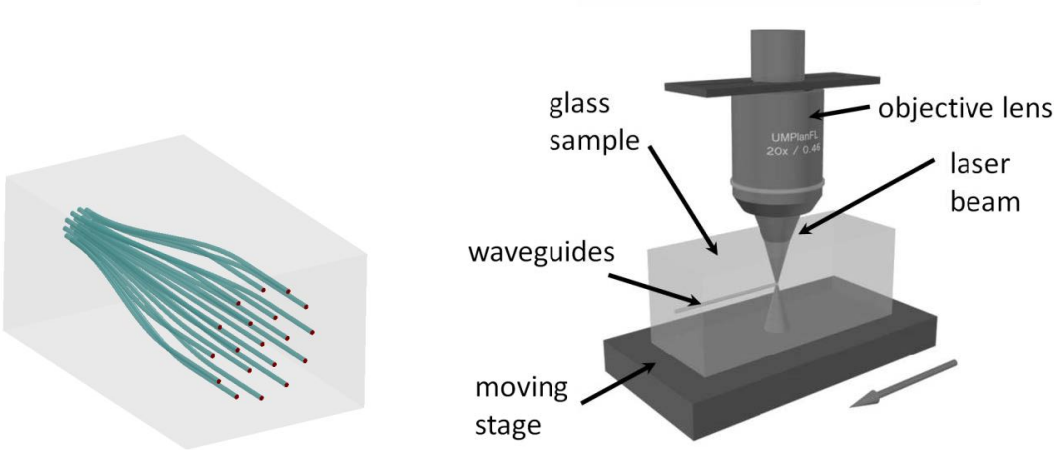


Figure 2.2: Four different ways of manufacturing a photonic lantern: A - Tapered MCFs; B - fused SMFs; C - ULI on a solid chip; D - An example of lantern made using B. The inserted panel shows the chip of ULI lantern. This figure is reproduced from [46].



(a) A photonic lantern on a chip. **(b) Manufacture rig using ULI technique.**

Figure 2.3: PL manufacture using the ULI technique: this figure is reproduced from [47].

achieved.

The adiabatic transition of MM to SM inside a PL can be explained using an analogy with the Kronig-Penney model in quantum mechanics (QM) [48]. As QM dictates that at the atomic scale, electrons display non-localisation and other wave-like properties. For simplicity, we can consider a system with two 1-D potentials as the two separate SM WGs in a PL transition. Therefore, there are two 1-D quantum potential wells in series within a bigger 1-D potential well for 2 electrons as shown in Fig. 2.5(d), the 2 electrons' wave equations interfere and tunnel through the smaller potential wells when the wells are very small and close to each other due to quantum tunnelling effect, while the big well's presence performs the major influence on the electrons' wave functions so the two electrons occupy two distinct energy levels allowed in the big well as if ignoring the existence of the two smaller wells. As the two smaller wells increase in size and distance from each other, the two electrons now experience the effect of the smaller potential wells. The electrons' wave equations meet the boundary condition in the smaller potential wells respectively until their energy wave equations can no longer leak through the potential walls of the smaller wells as the amplitudes of the wave functions increase and the energy decrease. The two electrons eventually decouple and reside in the lowest energy state in each respective smaller well as seen at the top of Fig. 2.5(d).

This is an analogue to the guided modes in PLs. The PL can be treated as a bundle with N number of very closely placed, identical SM cores within a larger cladding at the MMF end. At the small diameter and close proximity to each other, the incoming light can not be contained in any of the individual SM cores, which makes the individual SM core ineffective as WGs. However, the array of the SM cores collectively behaves as the core of a larger WG to the incoming light complete with the larger cladding around the composite WG. Regarding a guided mode in a 1-D SM WG, there are two components, the transverse wavenumber K_T and the propagation constant β, while $\beta = Kn_{eff}$ with K is the wavenumber and n_{eff} is a modal effective index. For a MM WG, the MM wavefunctions are the composite

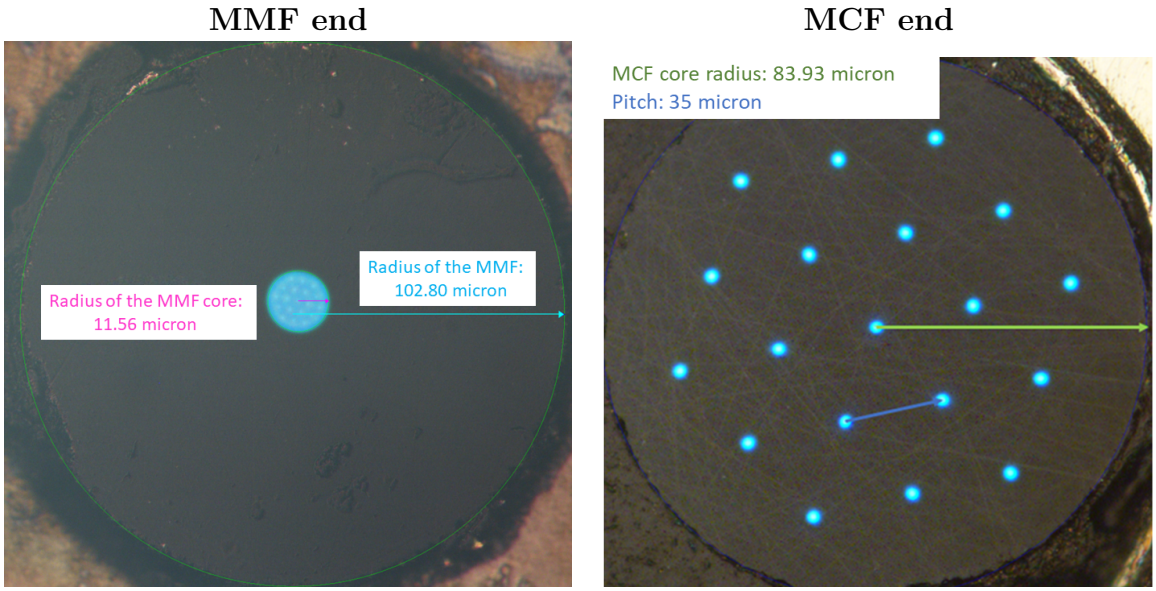


Figure 2.4: Photos of the MCF PL MMF and the SM MCF ends: The back illuminated MMF end (left). The illuminated MCF end shows the 19 individual SM cores (right).

functions with higher and lower n_{eff}. Hence the overall structure essentially is a MM WG supporting N number of non-degenerated super modes due to the homogeneous nature of the SM cores. The relationship between the two can be explained in Eq. 2.1.

$$K_n = \sqrt{{K_T}^2 + \beta^2} \tag{2.1}$$

As the diameter of each SM core gradually increases, the WG modes show a decrease in K_T and an increase in n_{eff}. For a given incoming light of λ wavelength, once at a certain core diameter, the nondegenerate super-modes start to be absorbed and trapped into the cores of each SM core. Mode beating occurs as a result of the coupling process, while each mode is still able to leak out of its core and interfere with adjacent cores at close proximity. The gradual increase in the thickness of the cladding around the individual SM cores resulting in the coupling distance to increase to the point that coupling between the cores no longer exists. The value of n_{eff} reaches the maximum and K_T decreases to its minimum when the diameter of the tapering region continues to increase, and these conditions meet the fundamental mode guiding conditions of the SM core. Then N number of non-degenerate MMs are now supported in each SM core as N number of degenerated fundamental modes, providing that the number of supported MMs at the MMF end matches the number of SM WGs at the end [36, 40, 43]. The wavefunctions of the original MM are still contained within the PL at the SM end as subsets of the original wave functions, hence in theory, the MM wavefunctions at the MMF end can be reconstructed by analysing its subsets at the SM end. The performance of a PL is also wavelength dependent due to the high efficiencies of the MM to SM conversion [36, 38].

Several types of PL have since been developed to meet particular desired requirements and applications [40]. In this thesis, the purpose of the work is to understand

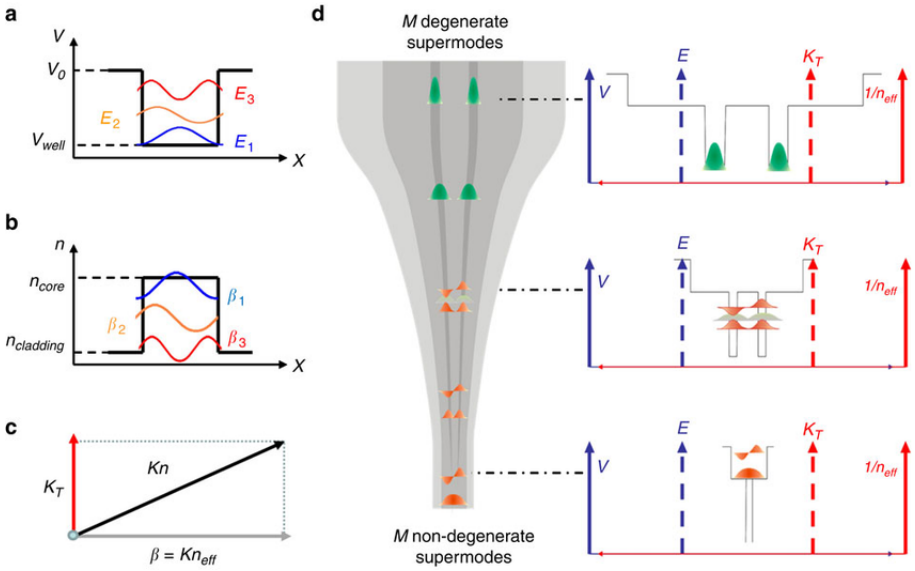


Figure 2.5: Quantum analogue of how a PL works: a is the electron in a 1-D potential well with different potential energy levels, E, that allows electrons to exist. b is the guided modes with different guiding constants, β, in a 1-D WG. This figure is reproduced from [49]

and exploit the complex spatial modes PL relationship in an AO system for analysis. In our work we will focus on a 19 core MCF PL (as shown in Fig. 2.4). A detailed analysis of the mode evolution of PLs can be found in [36, 38], however, we will not go into details of PL mode evolution for the purpose of this thesis.

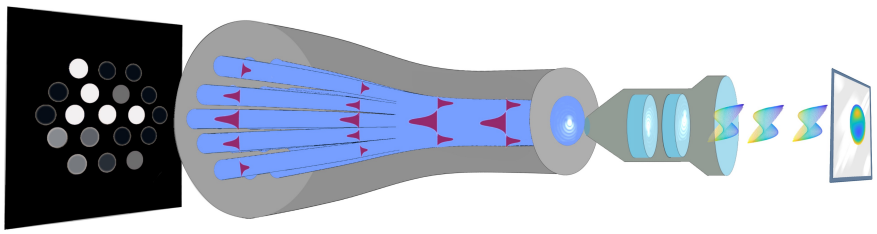


Figure 2.6: Wavefront propagation in PL: This is one of the configurations of the PL with 19 SM output as the MCF end, on the left side; the MM input that sustains 19 modes, as known as MMF end, is on the right. The traversing modes are coupled into the MMF end and decoupled at the MCF end into separated modes without much loss due to the gradual transition. The figure is not to scale

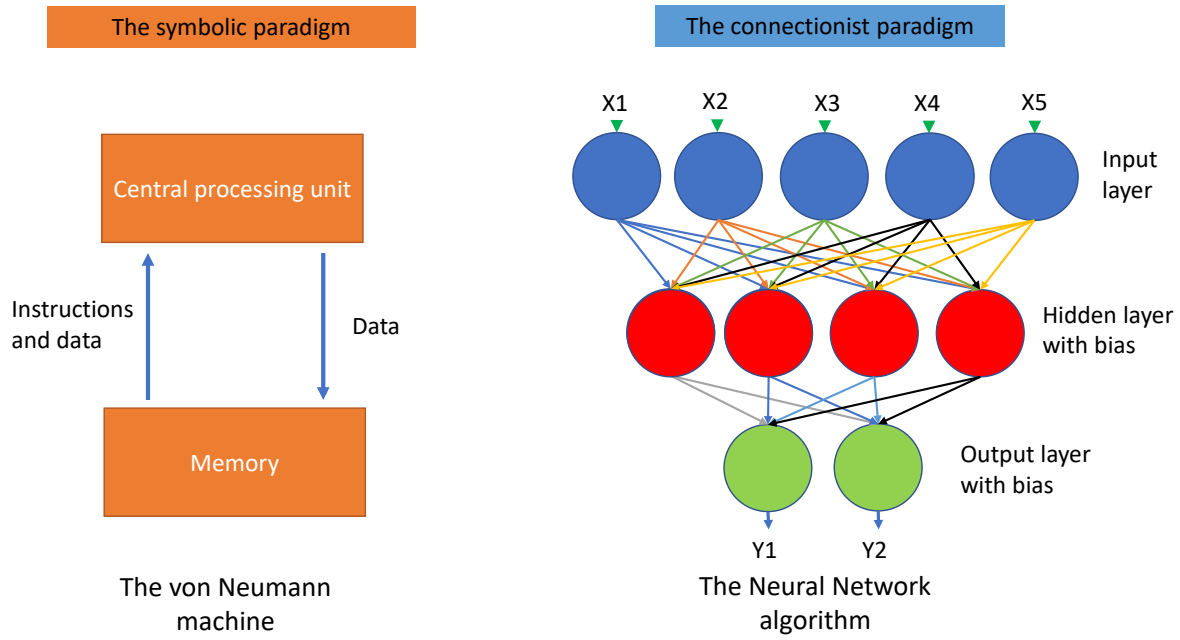


Figure 2.7: Comparison between the symbolic and connectionist paradigm: The left is an example of the symbolic AI stream: the von Neumann machine [51]. The right is an example of the connectionist: a NN algorithm with 1 hidden layer of different weights and biases. The circles represent individual 'neurons', the base processing unit in a NN [23, 52, 53]. Note: the different colours of the arrows represents a different value of weights embedded in the algorithm to lead to the next layers.

2.4 Deep Learning in Neural Network

Artificial intelligence (AI), in its general sense, has been successfully used in modern times in many areas of life [23] since the creation of the first computer in the early 40s. The quest to explore the intelligence of a human-made machine that can achieve complex tasks such as perception and understanding of the environment from its input; providing perception based on a common sense approach is always under consideration in real-life applications, as stated in [23]. There are two streams of research in AI, symbolic and connectionist [23]. Each of them has a very different way of solving problems, which is described in Fig. 2.7.

The universal Turing machine (UTM) [50], a hypothetical machine, which was described in a published paper by Ref.-[50] in 1936, is probably the best example of the symbolic paradigm and defines the computability of a problem using a set of well-constrained rules [23, 50]. All general-purpose modern computers are equivalently based on this model but work more akin to another type of symbolic model, the von Neumann machine [23], developed by Ref.- [51] in the 1940s. The basic workflow of the von Neuman machine can be described as shown on the left of Fig. 2.7. Although it was not a fully fledged AI algorithm, modern computer programming

is capable of understanding symbols and manipulating them without the computer itself knowing the semantics stored in their memory in the bit form [23]. Both Turing and von Neuman's model on the symbolic stream in AI has been used to develop the modern AI engine and found in many applications in different areas of industries based on the argument that all knowledge can be formalised and the mind or device can execute knowledge using sets of rules [23].

It was soon realised that the symbolic AI stream, or a so-called expert system, was not enough to solve things that were initially promised, such as producing general intelligence [54]. For example, AI in language translation was criticised by Ref.-[54] in his book that despite 10 years of research and millions of dollars, the machine translation system was still not operational [54]. Pattern recognition was particularly difficult for the symbolic AI stream due to the inability of a machine to create arbitrariness on its own given a limited set of examples [54]. Ref.-[54] concluded that in order to perform tasks such as pattern recognition and language processing, a new form of computer should be programmed based on a different theory from the symbolic paradigm [54]. This was also mentioned by Ref.-[55] in 1988, that symbolic AI had been generally successful over the years, but was insufficient in tasks such as chess playing, self-driving cars, language translation, etc.[55]. He emphasised a different approach to AI, the connectionist paradigm, for future AI development. Ref.-[55] proposed that the concept of connectionism was the answer to performing complicated computations in certain tasks such as recognising a picture [55].

The connectionist paradigm, which stems from the study of a biological brain[23, 53], was a parallel stream to the symbolic paradigm that can be traced back to the 1940s [23, 52]. The neural network (NN) is an example of the connectionist paradigm, which is an interconnected assembly of simple processing units, whose action is loosely based on an animal neuron cell which (in that case) performs the simple task of firing or not firing [23, 31, 55, 56]. The actual processing ability is contained in its connection with millions of other neurons as connection strengths, which are called "weights" , so the network as a whole causes the phenomena of "learning" in a machine via a process of adaptation from a set of training patterns given to the NN [23, 31, 52].

In animal neurons, the individual neuron achieves its function by firing or not firing a signal depending on the input strength [23, 53], for example, an animal's decision to attack or flee when threatened, which is said to be a basic cognitive function [53]. A set of input neurons get activated by an outside stimulus from the environment, and in the next layers of neurons, some are selectively activated through previously stored weighted connections from the input neurons, yet some are inhibited by the same inputs. This algorithm produces an outcome and rewrites some of the weighted connections depending on the new feedback, which is then stored in the NN: hence, the action of learning completed. This is a more mathematical term, is called credit assignment, an algorithm to find the weights in between the NN that make the NN produce a desired outcome [53].

Deep learning (DL) is a form of fundamental credit assignment, also known as learning in an artificial learning system or machine learning (ML), a subset of AI. DL gives a method for universal problem solvers with a time-optimal scale [52].

DL algorithms are widely used in the modern era for various applications, such as computer vision, language translation, and language processing [57]. The foundation of the DL algorithm is in the theory of artificial neural networks (ANNs) [23, 53, 57] in the search of artificial learning in machines. ANNs act like a set of decentralised neuron cells in an animal, utilising a series of multiple inputs to produce a single output. The output is made by categorising its input patterns, hence reducing the complexity of the algorithm as a whole for individual neurons. ANNs achieve this model by using a weighted sum of the inputs to lead to an output nonlinearly in the most common case [31]. This is analogous to an animal brain using the previous stored 'knowledge' of its experienced examples and interprets intuitively the new input by using a series of weighted distribution of actions. This process is very different from the symbolic AI approach, in that a computer has to be programmed in such a way that all the representations of actions are stored in a memory bank [23], and the symbolic AI approach can not learn from the new input from its intuition. As an example of the difference between the ML algorithm and the von Neumann machine, the ML is able to play a game of chess by learning how the pattern of each move can lead to a possible driven outcome and responding accordingly, while the other approach purely relies on the stored memory of the combination of moves leading to every single outcome there is in a chess game. However, any new combination will lead to a failure in response in the von Newmann machine, and hence it runs into storage issues very quickly in dealing with a problem that produces complicated patterns.

Artificial neural network(ANN), as stated above, are a computational algorithm that uses connected processors, called neurons, to produce a sequence of real-valued activation [52]. The earlier ANN construct is described theoretically in the paper by Ref.-[56] in 1943, in which they simply called the construct 'Nets' using the neurophysiological idea of all or none activation in animal neurons, also called the threshold logical unit (TLU). However, the net construct did not learn as modern ANNs do. Despite the net's inability to learn, their work became the foundation of the future development of ANNs, especially the TLU model of the ANN [23], and Rosenblatt in 1958 pushed forward the development of ANNs based on McCulloch and Pitts's work by introducing weights in the equation [58]; hence, the network was able to learn via learning curves. The TLU operates in binary: it takes in only two values, for example, 0 or 1. Each neuron is activated according to the credit assignment process via their statistical weights, obtained by adaptation characteristics from a set of training patterns [23]. It is the computer's attempt to mimic a biological brain and its learning process using a mathematical approach, linear regression [52], and the basic function can be seen in Eq. 2.2.

$$\sum_{i=1}^{m} w_i x_i + bias = w_1 x_1 + w_2 x_2 + w_3 x_3 + bias \tag{2.2}$$

$$f(x) = 1, \; if \; \sum_{i=1}^{m} w_i x_i + bias >= 0 \tag{2.3}$$

$$f(x) = 0, \; if \; \sum_{i=1}^{m} w_i x_i + bias < 0, \tag{2.4}$$

where w is the weight, x_i is each input in vector form, bias is the bias or threshold in the system and $f(x)$ is the outcome for the activation function. This is the most basic form of an ANN [52]. For example, if the input variables x_i have assigned values, such as $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, with the corresponding weight values, $w_1 = 0.3$, $w_2 = -0.1$, $w_3 = 0.7$ and evaluated as per Eq. 2.2 with bias given the value of -0.2, the sum is 1 which was greater than the threshold; therefore, the neuron gives an output, $f(x)$, of 1 from the activation function, seen in Eq 2.3. If the output is smaller than the threshold, then the output is assigned as 0, seen in Eq 2.4, which means that the activation function is inhibitory. In modern neural networks, such as that used in this project, these binary units are replaced with continuous-value neurons, which use a non-linear function (an activation function) rather than a hard threshold to pass a real-valued floating point number to connected neurons.

Backpropagation is a training algorithm to adjust the connection weights between each of the individual neurons [59–61], the detailed computation flow chart can be seen in Fig. 2.8; Backpropagation is essential for ANNs to learn from given examples [23] as it basically forms a feedback loop of the error report and adjusts the connection weight in the hidden layers according to the error report by comparing a desired output against the generated output in a descending gradient of the loss function [59]. Backpropagation was first theorised by Ref.-[59] in his 1974 thesis in its application in ANNs [59]. It was later rediscovered by both Ref.-[60] and Ref.-[61] separately with slightly different methods in their researches, for the training of noncyclic ANNs, also known as feedforward ANNs. The method done in Ref.-[60] popularised back-propagation for people to solve neural net related problems and renewed the enthusiasm in research into DL; while before that, the symbolic stream was still the more favourable research area for AI due to the time efficiency of the DL being low [23]. The more familiar construct of today's ANNs was established in Ref.-[61], by applying constraints in backpropagation, which allows for more generalised learning in ANNs with fewer free parameters. As a result, their network was proven to be applicable to a more real-life problem: successfully recognising a handwritten zip code from the US Postal Service using a relatively short time and a limited data set [61]. Backpropagation is often misunderstood as the whole learning algorithm for multilayer ANS [62] as it is actually the algorithm that computes the gradient rather than other algorithms, such as stochastic gradient descent, which is the algorithm that uses the gradient for actual learning. One of the drawbacks of backpropagation is the time involved in training ANNs on real-life problems [23], which was improved by Ref.-[63] in 1995 by developing ways to adapt a global learning rate to reduce training time [63].

The complexity of such a learning process using ANNs depends on the number of layers of neurons a particular NN contains [23, 52]. A single layer of neurons, without an activation function, can be used to classify two linearly separable classes, and thus is able to perform a calculation of a linear relationship between these two classes. Nonlinearly separable classes require more layers of neurons to perform and achieve the same purpose [23], since most of nature has nonlinearly separated classes. The number of layers of neurons is described as the "depth" of the NN. The first nonlinear NN algorithms to solve nonlinear problems in this way (learning from labelled examples) were developed back in the 1960s and 1970s [52], and were

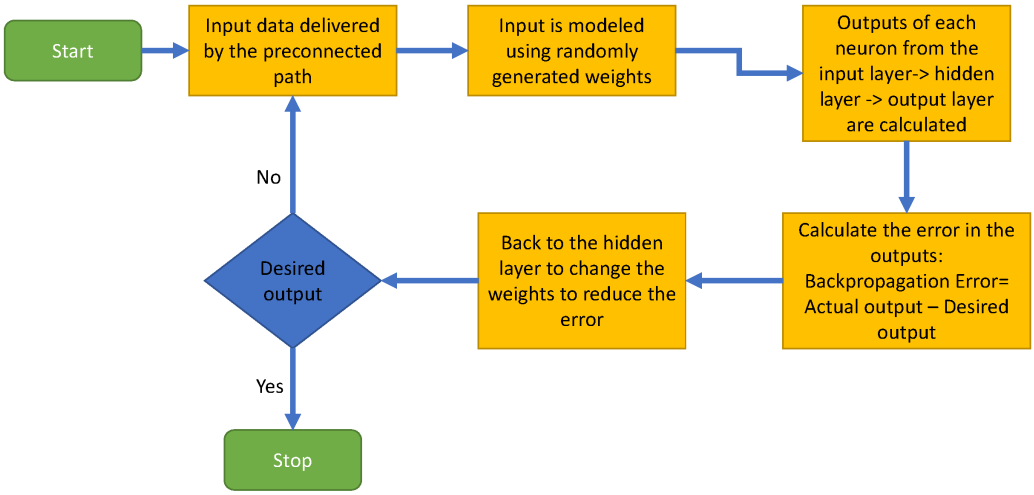


Figure 2.8: A program flow chart for backpropagation algorithm

supervised learning (SL) methods. A shallow NN that is discrete and differentiable in nature with arbitrary depth often employs teacher-based SL methods, such as backpropagation, which has been very effective [52]. Unsupervised learning (UL) on the other hand does not require labelled examples. UL has also been improved in recent decades in combination with SL to optimise its utility [23, 52].UL is used to uncover and disentangle hidden underlying sources of signals [52].

The ANNs used in this paper are an example of supervised learning (SL). The SL NN was chosen as the appropriate method due to the need to characterise the input phase map (using the Zernike modes) and the output fluxes, and recover a mathematical solution between the two. There is precedence of using SL NNs as a tool to reconstruct the PSF of the focal plane images with aberrations. For example, Ref.-[64] stated that NN as a real-time method to resolve focal plane images was possible, if the said NN was running on a high-performance GPU equipped with tensor cores. The results stated in Ref.-[65] demonstrated the advantages of employing a NN with a pyramid WFS.

In this paper, a similar use of SL NN is implemented in combination with a PL as the WFS to analyse the input phase information. The NN's role in the setup is to establish the relationship between the input phase and the fluxes in each of the 19 MCF end single-mode outputs. One of the issues of developing an optimised NN is to avoid overfitting of the model, which comes from sampling noise [66]. This leads to a model that does not generalise and thus can not predict with high accuracy. Therefore, many methods are used to reduce overfitting, such as introducing weight penalties, including L1 and L2 regularisation [64], when the validation performance deteriorates during the learning process. These are a type of hyperparameters, parameters for constructing or specifying the learning algorithm of a ML model [57]. The depth and amount of the hidden layers are also a type of hyperparameter, as well as the number of neurons contained in each layer.

Dropout is also a hyper-parameter to address both conservation of computation power and over-fitting when given a limited data set [66], it achieves the above by removing percentages of neurons [57] in each iteration of the learning process. The number of epochs (a complete run through all the training data) also is a hyper-parameter, which determines the time of the learning process and can also be fine-tuned by stopping when the validation accuracy starts to decrease to avoid over-fitting [57]. Others like batch size, size of each processed sample 'batch' before updating the model, learning rate, and the parameter that determines each step size of learning, are also hyper-parameters to enable optimisation of a NN model [57].

In summary, the progress of AI has spanned over two centuries, and the advent of ANNs has uncovered numerous applications in modern society, serving as a valuable instrument for data analysis. Specifically, NNs have been employed within WFSs as an analytical tool owing to their robust pattern recognition capabilities. Additionally, the NN can be optimised through its hyperparameters to yield high-performance analysis. The type of NN utilised in this thesis belongs to the category of SL NNs for analysing the experimental data.

Chapter 3

Laboratory Testing Procedure

3.1 Laboratory Setup

To validate the efficacy of the photonic lantern (PL) WFS, a number of laboratory experiments were performed. A laboratory test-bed was constructed, which allowed a WF with an arbitrarily chosen phase structure (imposed by a spatial light modulator (SLM)) to be focused and injected into a PL's MM input. The test-bed was placed in a temperature-controlled room with a temperature variance of +/- 0.1° C peak to peak at an average of 17.3° C to reduce the effect of thermal expansion of optical components.

A bypass filter was utilised to selectively choose a 70-nm bandwidth of light as the source for the experiment. The light was injected into an optic fibre and connected to an OAP collimator (Thorlabs RC08FC-P01) and passed through a set of neutral density filters to prevent saturation. A 2mm pupil stop was placed behind the bypass filter in the beam path to eliminate the Gaussian wings; the reshaped beam went through the polariser (PBS202) and reached the surface of the SLM which imparts a phase pattern with a dynamic range between 0 to 3.5 π rad peak to valley(P-V). The beam was then redirected by a beam splitter (non-polarising, R:I 50:50) into two beams. One beam was directed into a 200 mm focal-length lens and focused onto the detector, CAM2 (FLIR Blackfly BFS-U3-13S2M), viewing the resulting PSF. The other beam was focused onto the MM end of the PL within a fibreport mount (0.25 NA, 16.5 mm focal length). A fraction of this beam was reflected by the surface of the PL MM end, allowing an image of the PL's MMF end to form via a 200 mm focal-length lens onto the detector, CAM3 (FLIR Chameleon3-CM3-U3-13Y3M), simultaneously as reference. This feature allowed for easier alignment of the system. The rest of the light was channelled into the PL, the SM outputs of which (within a MCF) were imaged onto the detector, CAM1 (FLIR Grasshopper3-GS3-U3-32C4M), using a third 200 mm focal length lens. In the second iteration of the setup, the imaging happened via a spectrograph (seen Fig. 3.3). The custom-made spectrograph contains a collimator, a dispersive prism and 2xlens for focusing onto CAM1. The collimator is composed by a 15 mm achromatic lens(Thorlab AC050-015-B) and a 30mm achromatic lens (Thorlabs AC127-030-B). The collimator has a focal length of 16 mm. 2x60 mm achromatic lens (AC254-060-B-ML) are used to focus the light onto the camera. The dispersive prism gives the spectra resolution:

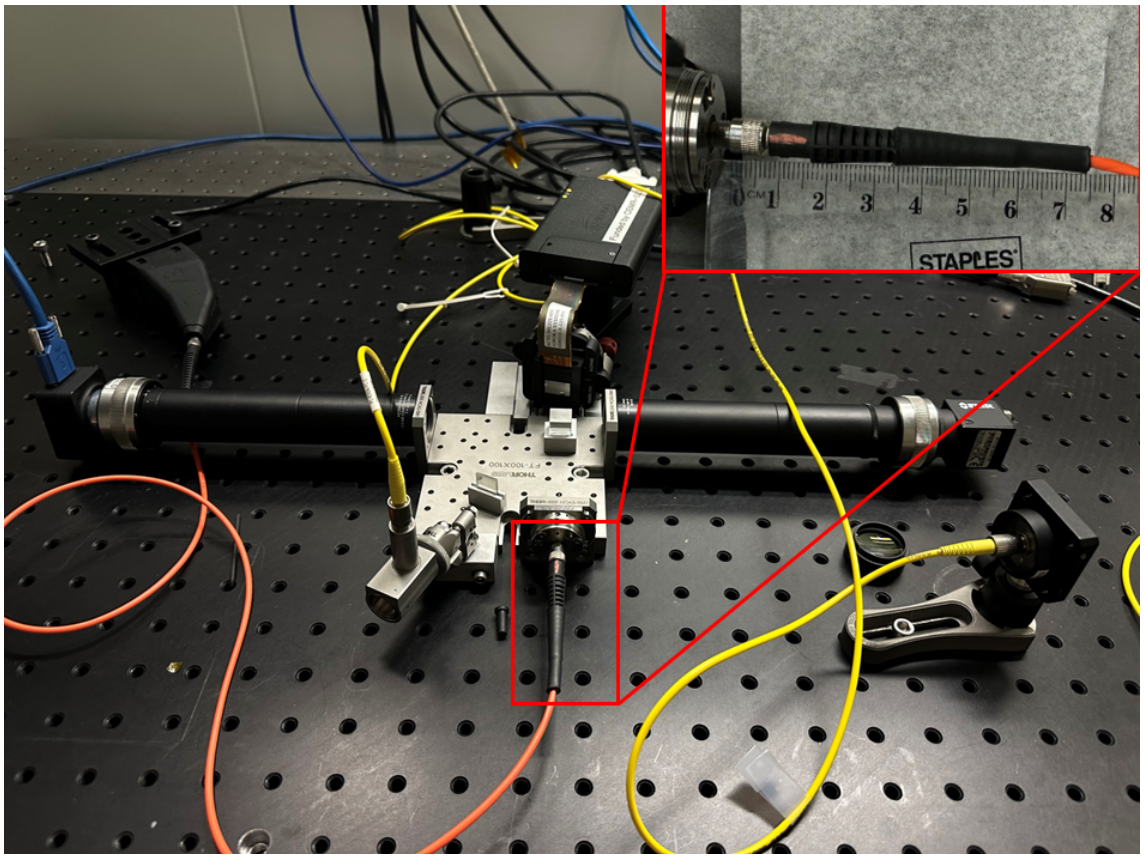


Figure 3.1: Photo of the experiment setup: this figure shows the in-lab test bed setup for the experiments. Noted that the PL is highlighted in the insert (≈ 7.4 cm in length). The rest of the length is purely for connection reach.The detailed configuration is depicted in Fig. 3.2

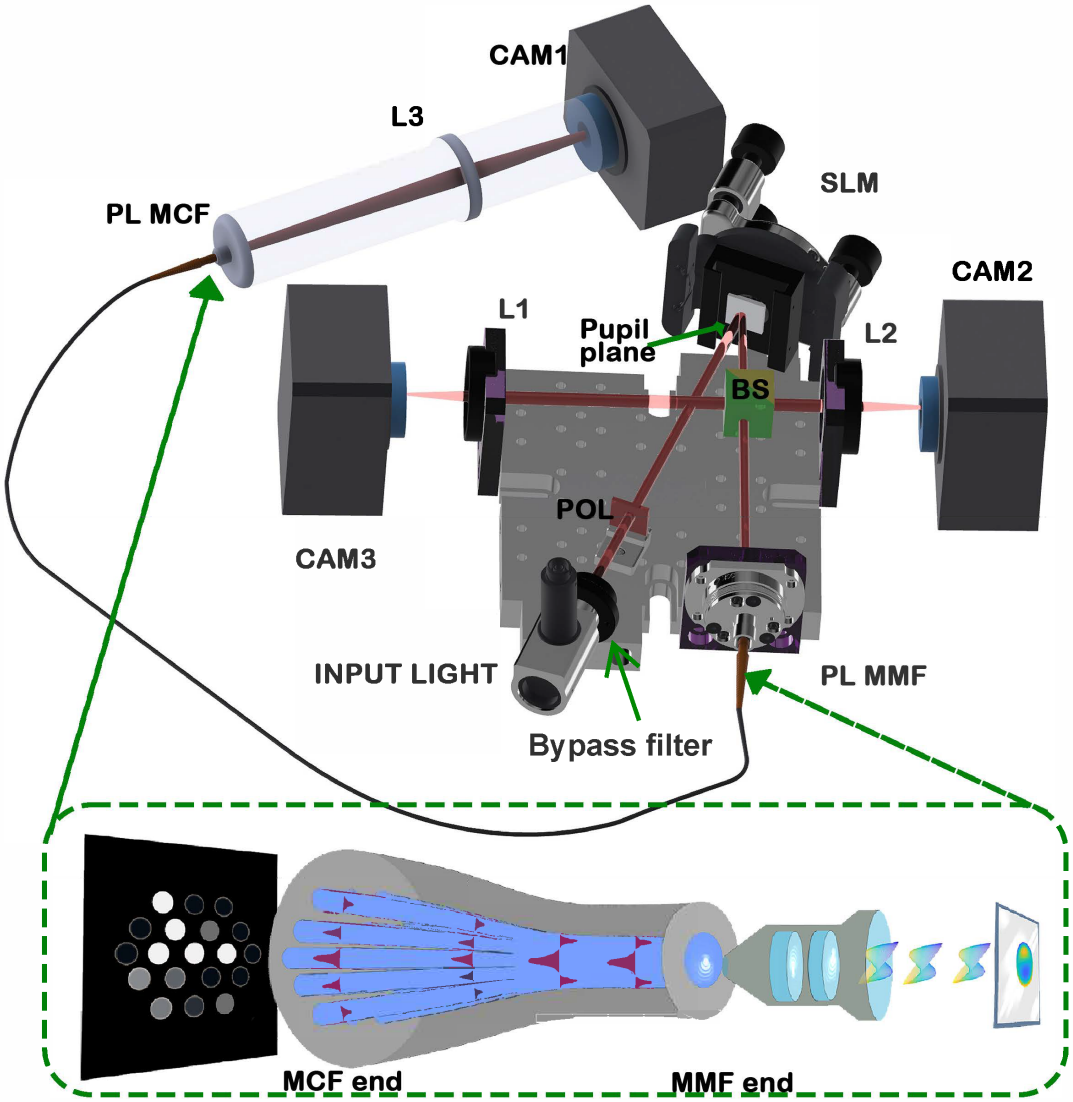


Figure 3.2: Experimental setup: A 70 nm bandwidth light (generated by SuperK Evo supercontinuum white light laser fitted with a 2mm pupil stop and a bypass filter centering at 704 nm) passes through a reflective collimator and polariser (POL) and onto the Spatial Light Modulator(SLM). The modulated light is then split into two beams of light by the 50/50 beam splitter (BS). Some light goes into camera 2 (CAM2) via a 200 mm focus lens (L2) to image the PSF; the other beam goes into a EFL=16.5 mm injection lens within a fibreport containing the PL's MM end. Camera 1 (CAM1) images the PL's SMF end via a 200 mm focal lens(L3). Some of the light beam is reflected from the MM core and directed towards CAM3 to monitor the image plane of the PL at the MM end via a 200 mm focal lens (L1).

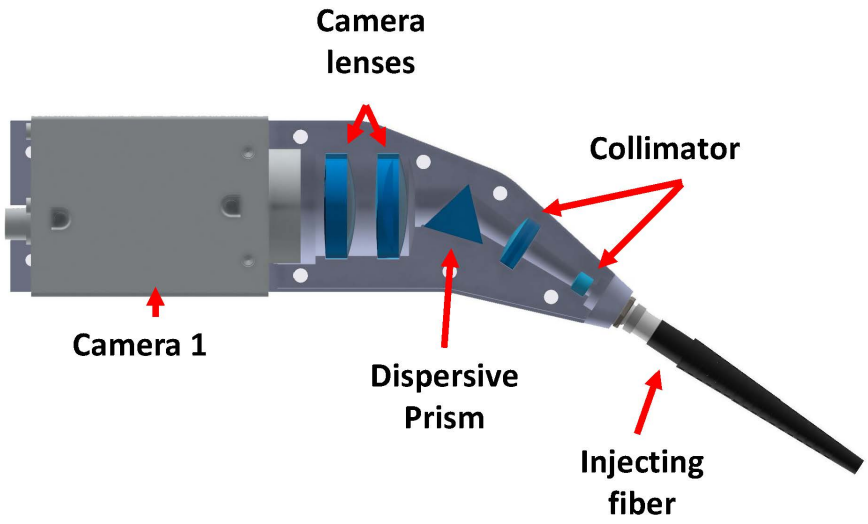
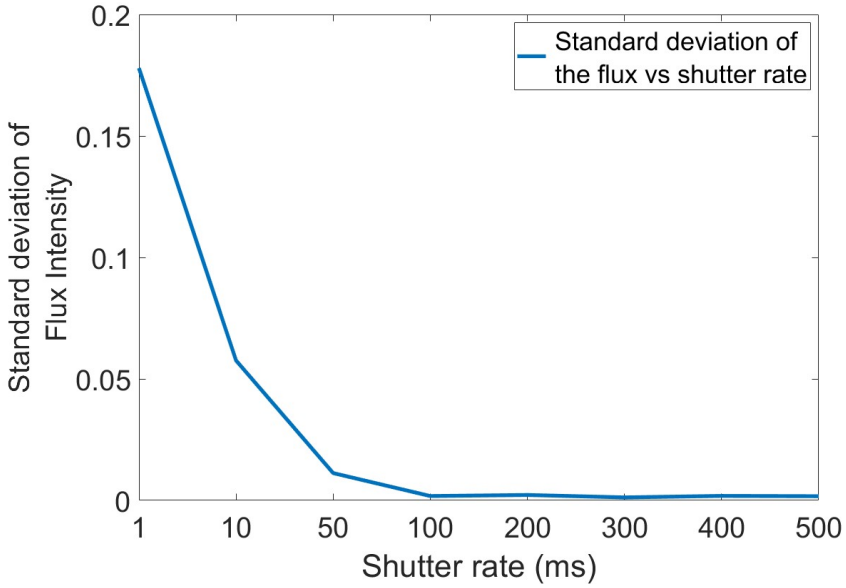


Figure 3.3: A schematic of spectrograph: shows the custom made spectrograph using 3-D printing material as the main body housing attached with a SMA fibre adapter.

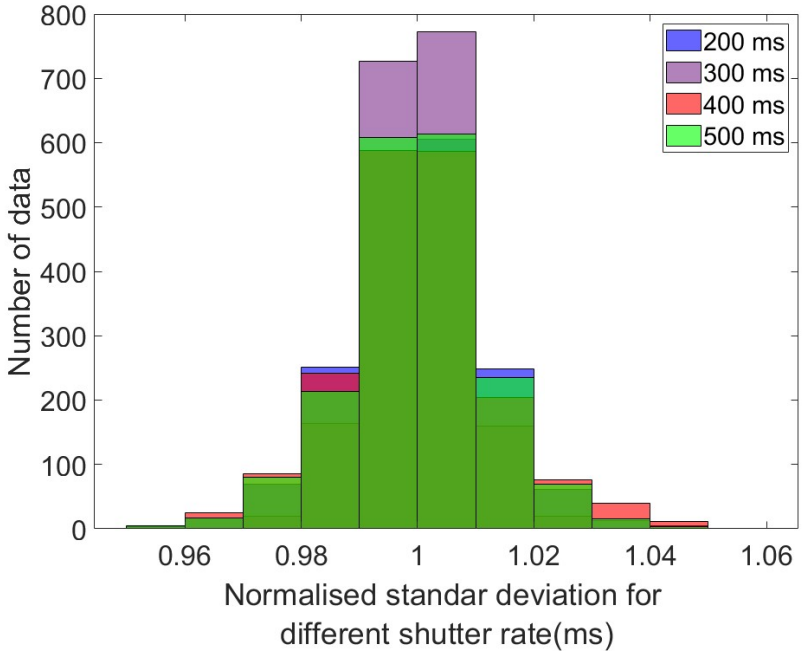
R=300 to allow each SM output to be measured as a function of wavelength.

The integration time of the cameras had to be chosen to minimise phase noise from the SLM. The Pluto Holoeye SLM has a refresh rate of 60 Hz, with modulation occurring at higher frequencies to produce intermediate 'gray' values. Thus the 3 cameras' integration times were set long enough to average out this phase noise, balanced against the need to acquire large volumes of training data rapidly. The global shutter rate was tested from 1ms to 500 ms, and the result is shown in Fig. 3.4.

The normalised standard deviation of the flux intensity tapering out as the shutter rate increased to 100 ms onward, as shown in Fig 3.4a. There is no statistical significance in choosing from 200 ms onward to minimise the phase noise as the remaining is most likely consisted of Cam 1's dark noise. However, another factor that accounts for the choice of shutter rate was the global flux intensity across three non-uniformed CCD cameras for visual purposes. As the result, 400 ms was chosen as the global shutter rate as it reduced the flux standard deviation to 1.84×10^{-2} flux intensity per ms, and a further increase in shutter rate induced over saturation on Cam 1 (see Fig. 3.4b). In addition, the process of displaying the phase map on the SLM is subject to a systematic delay caused by the refresh rate and software response. This delay can lead to errors in accurately injecting the correct phase information into the experiment setup. To mitigate this issue, a 0.1 second pause was incorporated into the acquisition program for each update of the new phase map. When collecting 40 data points, the overall acquisition time including built-in delays is 20 seconds. However, when factoring in the computer response time for processing and saving each individual file, the average time per file saved is 34 seconds. It translates to 18.9 hours of data collection per experiment of a total of 2000 data files with 40 data points.



(a) Shutter rate vs total flux stability: Normalised standard deviations of the total flux across PL 19 MCF output vs 8 different shutter rates of the camera setting.



(b) Total Flux Stability vs Shutter Rate from 200 to 500ms: Histogram shows the distribution of the normalised standard deviation of 19 flux intensity across PL 19 MCF outputs as camera shutter rate increases from 200ms to 500ms.

Figure 3.4: Two figures show the relationship between the total flux stability and the global shutter rate in milliseconds(ms).

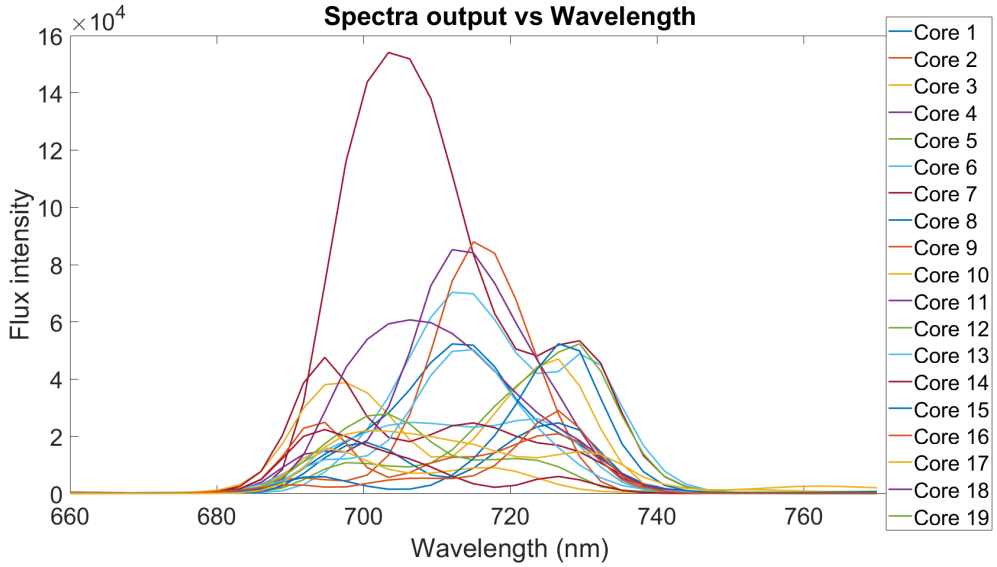
3.2 Wavelength resolved measurements

Existing WFSs in use are mainly limited to single-wavelength measurements [2, 8, 18]. Monochromatic source produces random speckle at image plane and makes detection of phase wrapping a major issue in bad seeing conditions. A WFS that is able to resolve measurements using polychromatic source is able to overcome these limitations; Thus, the test-bed included the low dispersion spectrograph. The bypass filter at the source (seen in Fig. 3.2 limited the wavelength range such that the spectra from each output core do not overlap (Fig.3.5).

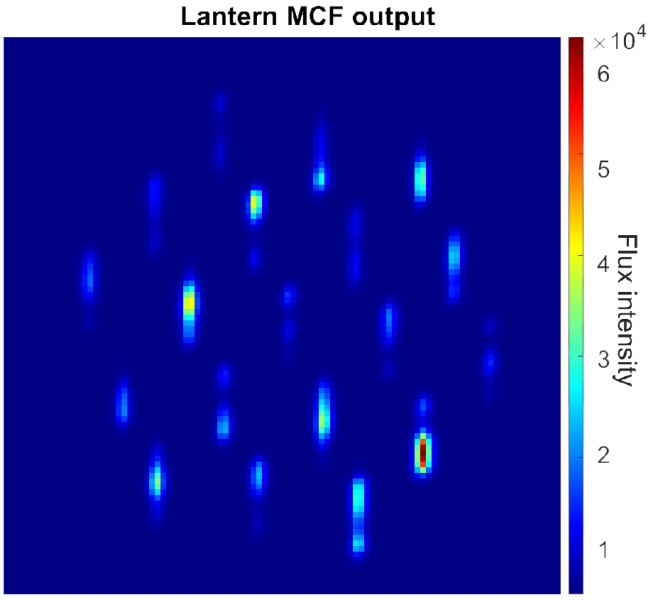
The spectrally-dispersed image of the MCF outputs was mapped to wavelength-calibrated fluxes by the following procedure ensuring the pixel-to-wavelength calibration of the resulting spectra is identical for all 19 output cores was important. For the initial wavelength calibration, a calibrated monochromator was placed at the light source to inject specific wavelengths. Four wavelength values ranging from 670 nm to 730 nm, with a step of 20 nm, were selected as reference points. These reference points were used to establish the interpolated equation for the wavelength-dependent intensity.The four sets of data were collected under the same condition to establish the relationship between the position of the 19 MCF cores and the correlated single-wavelength source. The four images were combined, as shown in Fig. 3.6b. The highlighted rectangular box in Fig. 3.6b shows the spectral extraction region for one of the output WG. The dimensions of the box are 2 pixels by 39 pixels, large enough to contain the entire spectrum across the 70 nm bandwidth. The 4 selected monochromatic wavelengths of the 19 outputs are enclosed inside a rectangular box (shown in Fig. 3.6a) as a circular pattern at the output. The pixel position of the peak intensity is extracted for each of the wavelengths at the 19 outputs. The box is divided vertically to get a step value of $\delta\lambda$, of 2.31 nm per pixel. The program, Matlab, used in processing matrix is in integer indexing, therefore an interpolated function was necessary to accurately describe the relationship between the flux centroid and the wavelength. Fig. 3.6a shows a 2-D fitted function over the 4 positions in Fig. 3.6b as each line indicates 1 fitting function for position vs. the wavelength for 3 outputs. It is important to note that none of the 19 cores contains the same fitting function, due to the slight positional difference on the camera across the 19 cores, seen in Fig. 3.6b.

3.3 Measurement procedure

A pattern of the designed modes was generated by manipulating indexes of a zero value matrix (181 x 181 in unit pixels) with randomly assigned coefficients with a predetermined combined wavefront RMS error range: 0.88 rad and 1.5 rad on average, respectively. The tests were conducted using different combinations of Zernike phase maps, including 9, 14 ,and 19. The Matlab function that was used to achieve this was Zernfun [67]. A phase map was made by linearly adding the desired modes together. The process involves injecting specific phase maps onto the SLM screen for manipulation, which then propagate through the PL and result in flux intensities across the 19 PL outputs. The data is acquired via CAM 1 as seen in Fig. 3.2.

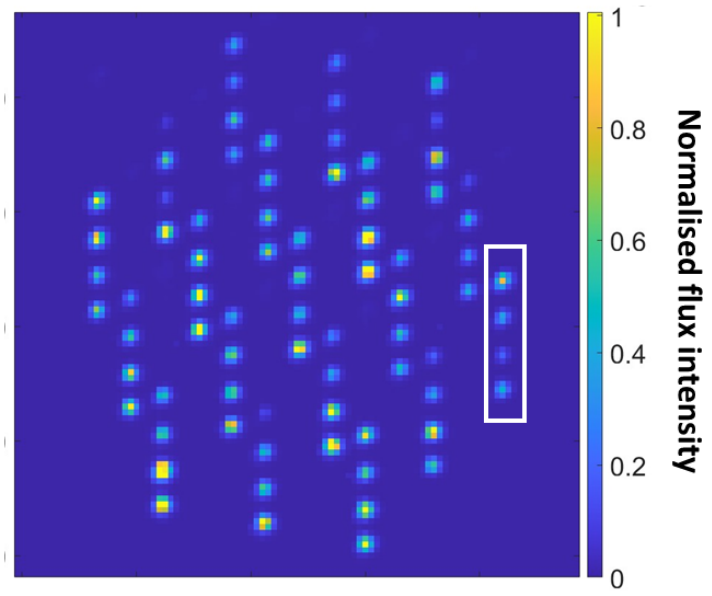


(a) Spectra output vs Wavelength: Plot of wavelength vs the intensity of the flux from the 19 MCF cores.

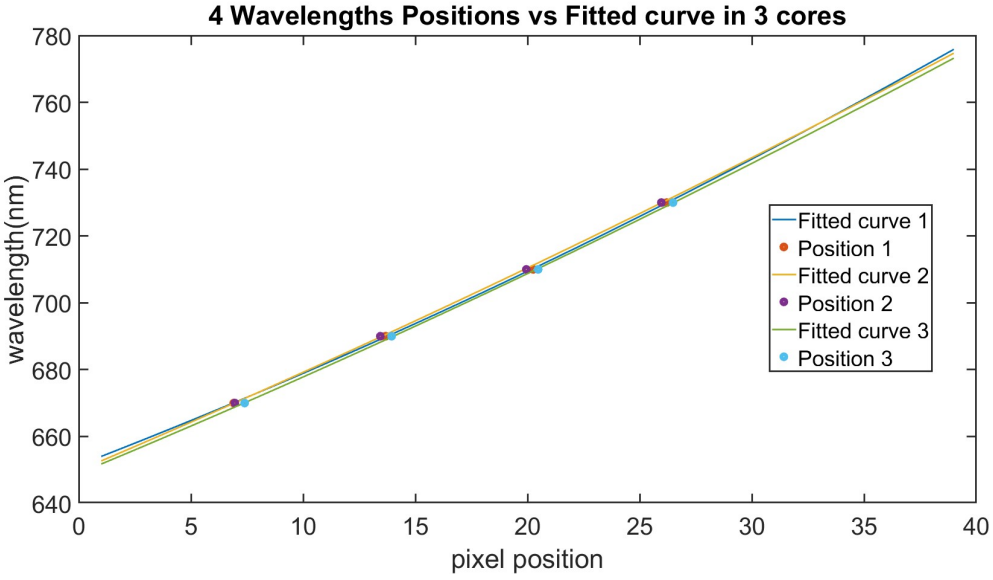


(b) Lantern Image Plane spectra output: The raw camera image of the same MCF cores as the data shown in Fig. 3.5a.

Figure 3.5: Wavelength-resolved output of the MCF PL illuminated by a light source with a wavelength from 660 to 780 nm. The total wavefront at the input is 1.5 radian RMS in average.



(a) Raw camera image of the 4 calibration wavelength positions for each of the 19 cores: The highlighted box represents 1 of the cores with 4 different wavelengths (670nm to 730 nm with a step of 20 nm, from top to bottom).



(b) Wavelength calibration curve: this graph shows 4 fitted lines to the 4 wavelength's pixel positions respectively. The location of each set of 4 points is the graph presentation of the same 4 dots in the rectangular box in Fig. 3.6a. Noted: only three out of 9 outputs are shown here for visual clarity.

Figure 3.6: Wavelength calibration: spectral extraction process used for the spectrally-dispersed MCF outputs.

Dark frames were produced for each camera by averaging 200 unilluminated images, which were subsequently subtracted form each measurement. An automated control loop, written in Matlab, was used to construct and display random phase maps on the SLM, and save the displayed phase maps with the corresponding camera images in one file. Due to the limitation of the computer's random access memory (RAM), a set of 40 data points was allowed to be stored in individual files before running out of RAM. For a full run of data acquisition, a total of 2000 such files were collected for an individual experiment. All raw data were stored as Matlab files; 9 different experiments were conducted for characterising the 19 MCF output PL as WFS.

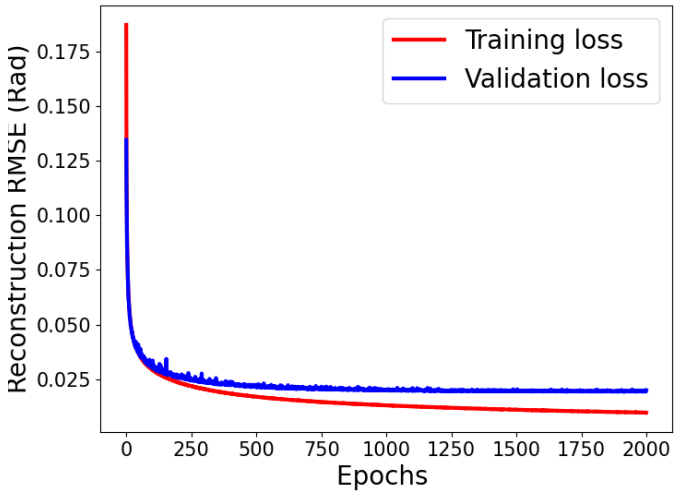
3.4 Interrogation of the NN structure for analysing the PL wavefront input

Some of the hyperparameters greatly affect the NN prediction while training the model; one of the hyperparameters is called dropout rate as mentioned previously in Section 2.4. The dropout rate is usually used to avoid overfitting the model, as overfitting undermines the ability of the model to predict unseen new data [66]. The different dropout rates were tested on the 9 Zernike phase map with on average 0.88 rad incident wavefront RMSE, the result can be seen in Table 3.1. It should be noted that when the dropout rate was set to 0, overfitting occurs in the model (shown in Fig. 3.7a, which was not an ideal situation either as it has been discussed in Section 2.4 earlier). The optimal amount of the dropout rate is at 0.1×10^{-1}, where the test and validation losses are overlapping, as seen in Fig. 3.7b. When the dropout was too large for the model (dropout rate = 0.3) the model became underfitting (shown in Fig. 3.7c). This is again not ideal as the algorithm for learning in the model has underfitted too much in each iteration of the learning process, so the test set under-represents the validation set, and this over-representation of the data is reflected in the loss curve since the training loss has higher reconstruction RMSE than the validation loss. Another hyper-parameter, epoch, was turned to optimise the NN performance. As seen in Fig. 3.8, the linear relationship between the epoch and the time, the amount used to train the model, presents a cost in the application of PL-NN methods for wavefront detection. As discussed previously in Chapter 1.3, the computation time is directly related to the temporary error, which impedes overall AO performance. However, the NN reconstruction RMS error demonstrates an inverse, non-linear relationship as there is a sharp decrease when the epoch increases from 100 to 500, shown in Fig. 3.8, and a much slower decreasing trend afterwards. It eventually gives no more benefits to extend the training time, the epoch, to decrease the reconstruction RMS error. The NN reconstruction error also depends on the NN structure in terms of linearity and the number of hidden layers applied when building a model, which can be explained by the complexity relationship mentioned in Section 2.4. However, the trade-off is time efficiency during training, as the deeper the NN, the more it penalises training time [57], a known shortfall for most DL models, including NN. Supervised NNs are also known to require a large amount of data while building an optimal model, as can be seen in

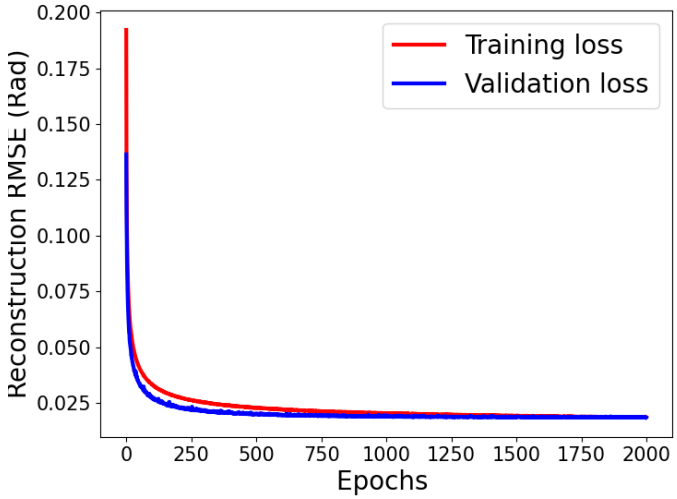
Fig. 3.9. As the larger the data set is given to the NN, the smaller the reconstruction error in RMS becomes, with almost a reverse exponential pattern. A three-layered NN may not necessarily be classified as a deep network; however, it still contains 3 non-linear hidden layers. The more Zernike terms included in the phase map, the more complicated the relationships between input and output, hence the deeper the NN required to deal with the increased variables to predict [66]. Currently, as in the DL NN used here, increasing the number of layers does not necessarily mean increased performance, as discussed in Section 2.4. This is also true for the number of data required to train the NN if the model is kept the same. Using only half of the 80000 data points, we were still able to predict with similar accuracy for 9, 14 and 19 Zernike terms in the smaller RMS WFE range at the input. However, the same can not be said with a larger incident RMS WFE for the same terms, which is possibly due to the current NN construct still not being optimised in dealing with the high nonlinearity of the input wavefront and also the fact that SLM displayed more phase wrapping at the range over its maximum P-V threshold, so it affects the NN's ability to accurately predict the coefficients.

Drop out rate	Reconstruction RMSE (rad)
0.00	1.90×10^{-2}
0.1×10^{-1}	1.84×10^{-2}
0.2×10^{-1}	1.84×10^{-2}
0.5×10^{-1}	1.92×10^{-2}
1.0×10^{-1}	2.30×10^{-2}
2.0×10^{-1}	2.80×10^{-2}
3.0×10^{-1}	3.40×10^{-2}

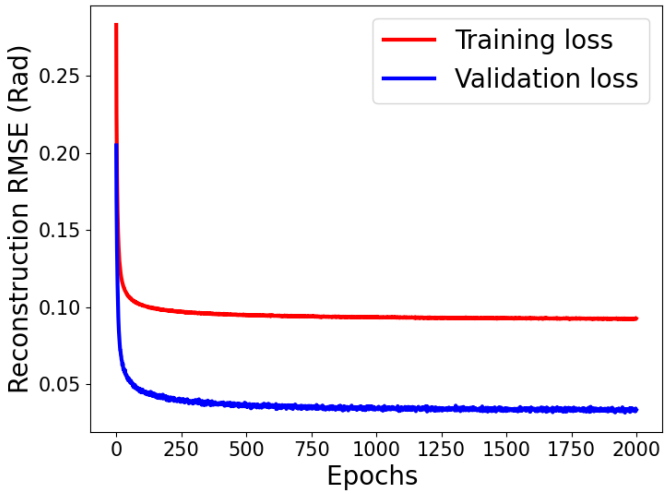
Table 3.1: NN hyperparameter: dropout rate vs the reconstruction error: different dropout rate and the reconstruction RMS error for 9 Zernike with input total wavefront RMSE of 0.88 rad.



(a) Dropout rate: 0: the training loss shows a lower value than the validation loss , indicating a overfitted model.



(b) Dropout rate: 0.01: the training loss overlaps with the validation loss , indicating a well-fitted model.



(c) Dropout rate: 0.3: the training loss tapered out at a higher value than the validation loss, indicating a underfitted model.

Figure 3.7: Loss curve with two values of dropout rate using the same NN construct.

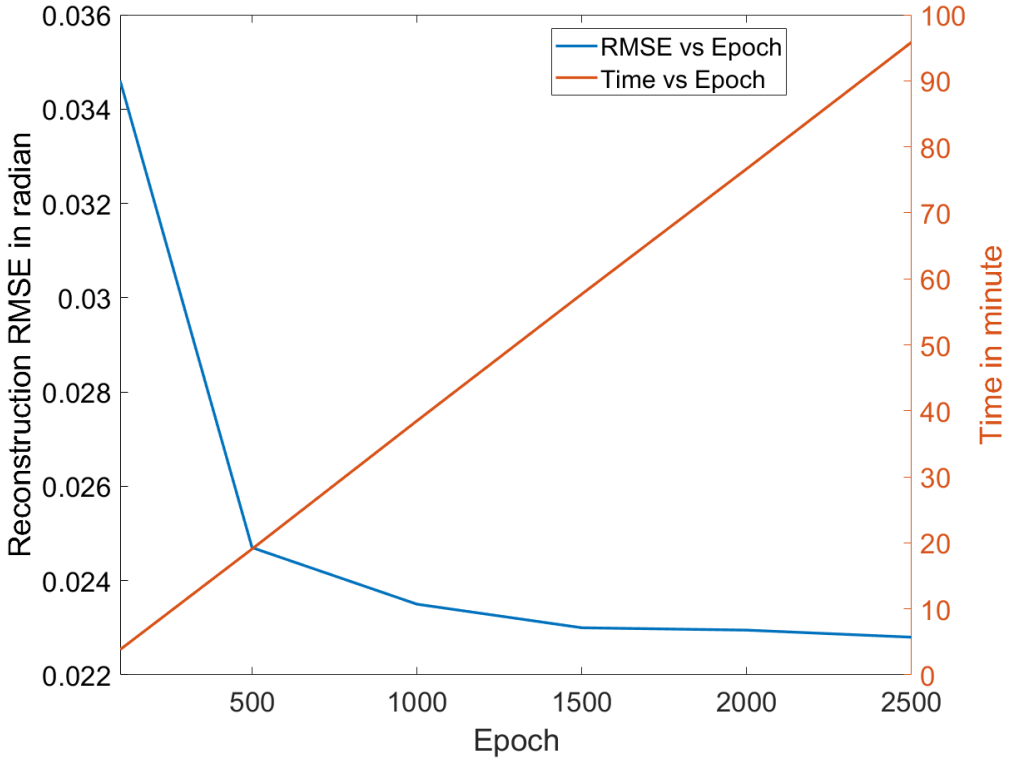


Figure 3.8: Number of Epoch vs time used to train NN: Relationship between the number of epochs used to train the NN model with the reconstruction RMSE and time consumed to build the model while the amount of training and validation data sets were unchanged

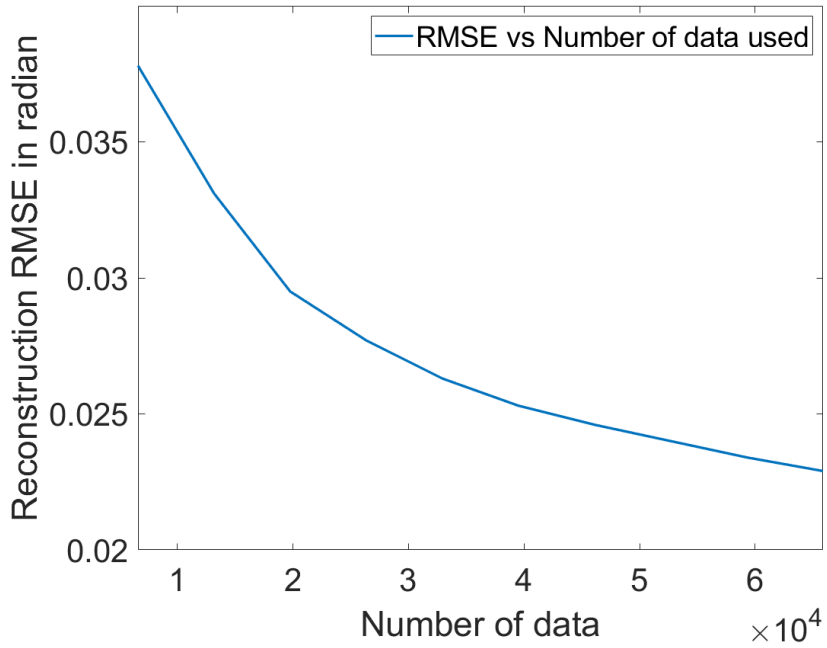


Figure 3.9: Number of data used in training the NN model vs reconstruction error: using the same hyperparameter with different number of data quantity the RMS reconstruction error

Chapter 4

Experiment result

To test the ability of the PL-WFS to accurately reconstruct the incident WF, a number of different laboratory tests were conducted. In each experiment, a different type and magnitude of WFE was imposed, and the reconstruction accuracy quantified. For a general test of low-order WF sensing, the Zernike polynomial was used, a commonly used basis for representing optical WF aberrations. Both the number of terms (corresponding to the number of spatial modes sensed) and the magnitude of the WFE was varied, see in Fig. 4.1. Next, the ability to sense the problematic low-wind effect modes (also called island modes)/ petal modes was tested, using a basis specifically designed to emulate these modes. Finally, the ability of the PL-WFS to reconstruct simulated atmospheric turbulence was tested. The performance of each test was evaluated based on the reconstruction RMSE, which is the RMSE of the residual map between the actually-applied and reconstructed WF RMS.

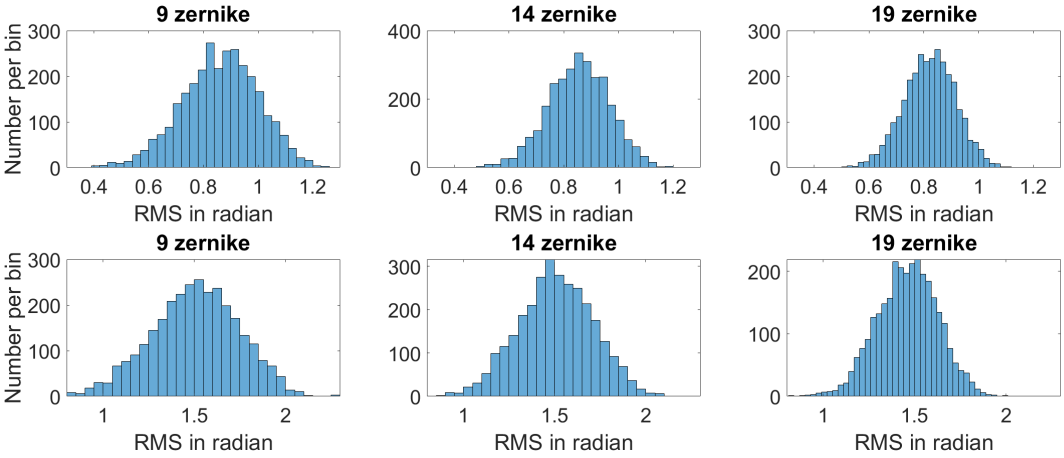


Figure 4.1: Distribution of total incident RMS WFE for the different Zernike experiments: the top row is for the low WFE phase maps; the bottom row is for the high WFE phase maps.

4.1 Analysis of wavelength binning

Before conducting all the experiments, the optimal bandpass needed to be established. Since the output is spectrally dispersed, it is possible to choose the centre wavelength and bandwidth of light used in the analysis by integrating over the chosen spectral channels. An optimal bandpass was chosen using the procedure described here in the experiments described above. This same bandpass was used for all experiments.

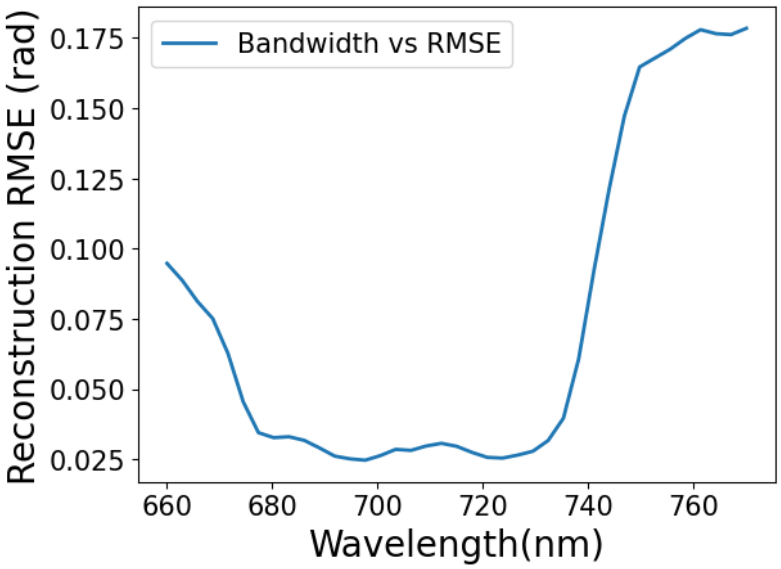
A test set consisting of 9 Zernike modes with an average incident RMS WFE of 0.88 rad was selected, with identical raw test data being utilized for all tests in this task. In the first experiment, after obtaining individual spectra from the PL output image as previously described, the resultant sets of 39 fluxes corresponding to 39 wavelength channels (ranging from 655.84 to 751.14 nm) were each used independently to train and test a neural network tasked with the reconstruction of the wavefront coefficients.

Fig. 4.2a illustrates that the most effective wavelength range for the PL WFS was between 680 nm and 730 nm. Notably, the poorest performance of the PL was observed at 765.84 nm (with the best RMSE of 1.75×10^{-1} rad) and the most effective at 693.17 nm (with the best RMSE of 2.50×10^{-2} rad) when comparing the reconstruction RMSE (see Figure 4.2a). This pattern was anticipated and can be largely attributed to the decrease in flux, resulting in a decreased signal-to-noise ratio (SNR) at the upper and lower limits of the bandpass filter positioned in front of the source.

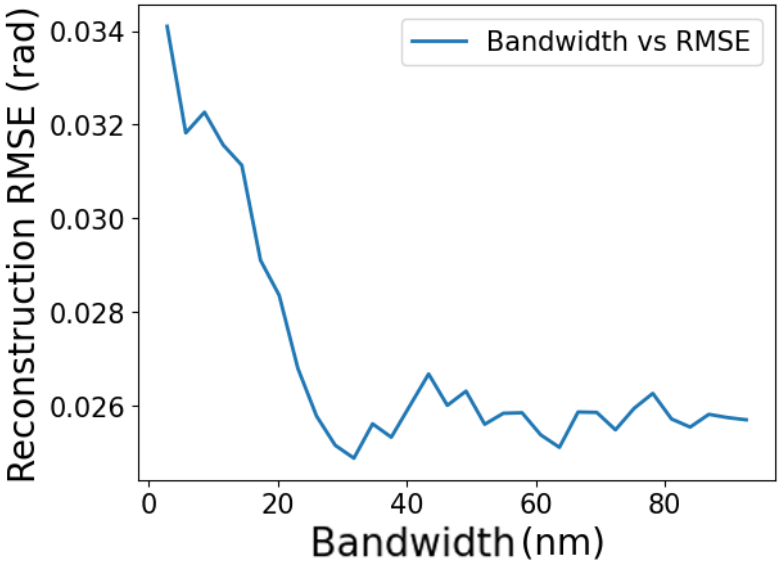
The measured fluxes were integrated across a range of different bandwidths, with the integrated fluxes for each bandpass choice subsequently used to train and test the network. The best performing bandwidth was the found, as shown in Figure 4.2b. The narrowest bandwidth of approximately 2 nm was the least effective due to low SNR, and a rise in reconstruction RMSE was also observed when larger bandwidths were employed. This is expected when the bandwidth becomes large enough that the flux at each wavelength is not linearly proportional to all other wavelengths in the bandpass, e.g. when a null is present at some wavelength. The most effective bandwidth was demonstrated to be 34 nm, which, with a centre wavelength of 680 nm, was adopted for all experiments presented here.

4.2 Low-order WFE with Zernike modes

Zernike polynomials [68] are widely used in AO as the basis for optical aberrations due to their orthogonal nature [2] and the fact that their lower-order series mathematically describe the classic optical aberrations such as tip, tilt, astigmatism, etc. [69]. This basis was thus used here to simulate general low order wavefront aberrations. An important question to be answered is how many spatial modes a PL-WFS can sense. Due to the available degrees of freedom, the number of output WGs places a hard upper limit on the number modes that can be sensed. But in practice, due to symmetries in the PL lantern and the mismatch between Zernike basis and the modes natively supported in the PL, this number would be expected to be lower. These experiments used sets of phase maps made with 9, 14 and 19 Zernike terms



(a) The best reconstruction RMSE for a single wavelength across a wavelength span between 660 nm to 768 nm.



(b) Bandwidth vs reconstruction error: The change in reconstruction RMSE versus the different integrated bandwidths used

Figure 4.2: The relationship between the wavelength and performance of the PL setup

(excluding the piston mode). For each of these, a set with lower WFE (0.88 rad RMS) and a set with high WFE (1.5 rad RMS) were tested.

In the first experiment, for each test measurement, the first 9 Zernike modes (excluding the piston mode) were linearly combined to form one phase map on the SLM. Each of the 9 Zernike terms was given a coefficient randomly chosen from a uniform distribution ranging from -0.26 to 0.26 rad in RMS WFE. The combined 9 Zernike modes yielded an average incident RMS WFE of 0.88 rad. An example measurement is shown in Fig. 4.3. For each applied phase map, the PL's output fluxes at its MCF end, the image of the PSF and the image of the back-reflection of the injection are saved.

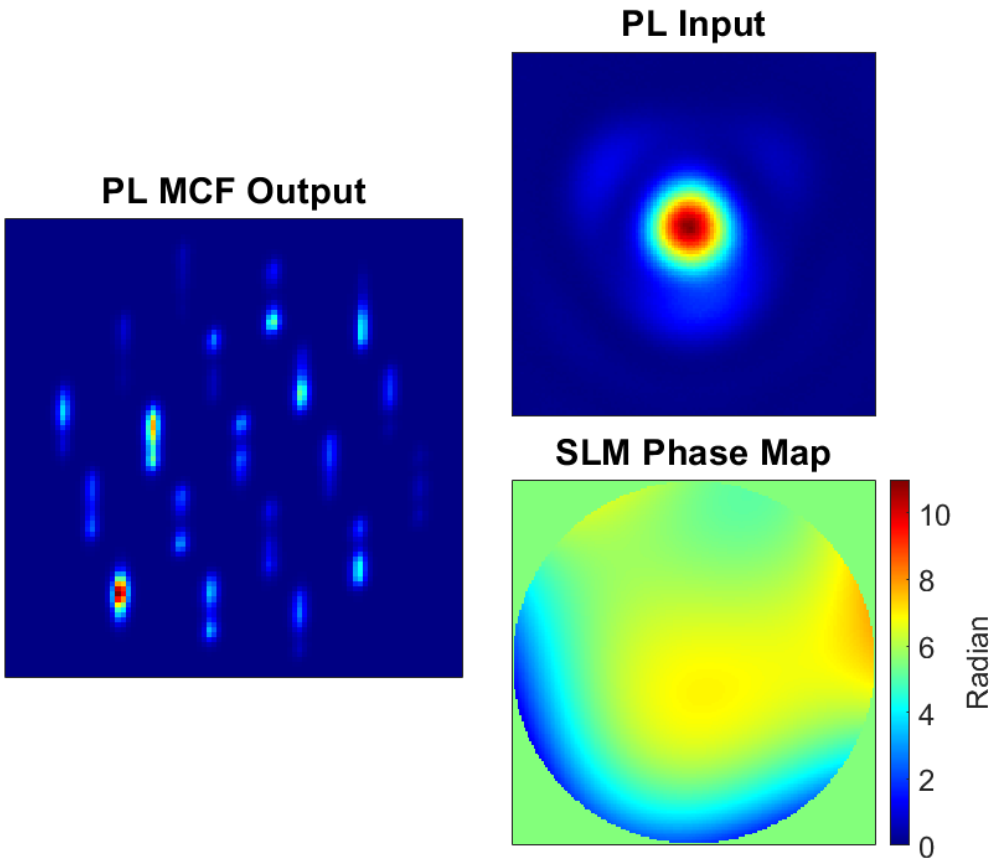


Figure 4.3: An example of a (polychromatic) individual measurement for the 9 Zernike data set. As with all other types of test data, random modal coefficients were chosen to produce a phase map describing low-order WFE, which is applied to the SLM (bottom-right). The resulting PSF (top-right) is injected into the PL, and the 19 outputs of its MMF end (seen here spectrally dispersed) are measured by a camera (left). This is repeated of order 10 000 times to produce the data set.

While many existing FP wavefront sensing algorithms rely on a linear approximation for the relationship between WFS signal and wavefront phase[8, 17, 20, 26, 30, 31, 69], in practice this is not always a good assumption as real-world WFE often have magnitudes placing them well outside the locally-linear regime. This is especially true for modes such as LWE. To test the ability of the PL WFS and neural

Number of Zernike Terms	Average RMS WFE applied (rad)	Number of neurons and NN layer	Epochs	Dropout rate	Learning rate
9	0.88	2000-1100-100-9	100	0.004	1.00×10^{-4}
9	1.50	2000-2000-200-9	100	0.05	1.00×10^{-4}
14	0.88	2000-2000-2000-14	100	0.22	1.00×10^{-4}
14	1.5	2000-2000-2000-14	300	0.40	1.00×10^{-4}
19	0.88	2000-2000-2000-19	200	0.40	1.00×10^{-4}
19	1.5	2000-2000-2000-19	200	0.40	1.00×10^{-4}

Table 4.1: Summary of neural network parameters

network to perform non-linear wavefront reconstruction, the total WFE were also constructed with an average RMS WFE of 1.5 rad. The total peak-to-valley (P-V) error in wavefront was found to be up to 11 rad on average, depending on the number of modes that were combined to create the final phase masks. This distribution of wavefront included many instances with very large RMS WFE (up to 2 rad), firmly in the non-linear regime - see the histogram in Fig. 4.1 and Fig. 4.18. For both WFE magnitudes, a set of 80000 measurements (consisting of random combinations of the 9 Zernike modes) was taken. Measurement sets using the first 14 and first 19 (non-piston) Zernike terms, for both RMS WFE magnitudes, were acquired in the same fashion. A NN was then used to test the performance of wavefront reconstruction for each of these experiments. In each case, 80% of the data was randomly selected to train the NN, and the other 20% kept separate and used to evaluate the performance of the trained network. Withholding a separate test set allows one to detect if overfitting has occurred, and proves that the network is able to generalise to new, previously unseen data. Fully connected NN were used, with the network parameters tuned for each case. The network parameters are summarised in Table 4.1. The NN was built in Tensorflow [70] using the Keras API [71], and trained on an NVIDIA GeForce RTX 2080ti GPU.

The reconstruction accuracy results for all Zernike experiments are given in Table 4.2. For the 9 Zernike term, with low incident WFE (0.88 rad RMS) set, the reconstruction RMSE per mode was 2.21×10^{-2} rad. The set of results from high incident WFE (1.5 rad RMS) shows an almost threefold increase in reconstruction error, with a value of 6.89×10^{-2} in rad. The resulting predicted and actual coefficients for the phase maps can be seen in Fig. 4.5 and Fig. 4.6 for two different incident WFE. The lower incident WFE (Fig. 4.5) gives a better prediction in coefficients across all 9 Zernike modes than the higher WFE data set (Fig. 4.6) in general. Similar differences are also observed in the 14 and 19 Zernike term sets, where the reconstruction RMSE per mode was 5.70×10^{-2} rad for Fig. 4.7 and 8.50×10^{-2} rad for Fig. 4.9 respectively at the low incident RMS WFE; the reconstruction error increased to 1.72×10^{-1} rad in Fig. 4.8 and 2.07×10^{-1} rad for Fig. 4.10. The more detailed breakdown of RMS WFE per mode are displayed in the figures from Fig. 4.5 to Fig. 4.10. In the best cases, the reconstruction residuals for each mode are barely

Number of Zernike Terms	Reconstruction RMSE per mode with 0.88 rad incident wavefront	Reconstruction RMSE per mode with 1.5 rad incident wavefront
9	2.21×10^{-2}	6.89×10^{-2}
14	5.70×10^{-2}	1.72×10^{-1}
19	8.50×10^{-2}	2.07×10^{-1}

Table 4.2: Summary of reconstruction performance, in terms of reconstruction RMSE, for all the Zernike-based experiments.

visible (coinciding with the 'Residual=0' : the black dashed lines) in each of the figures, showing a very high accuracy in predicting coefficients for the Zernike terms used. However as the magnitude of applied WFE and number of Zernike terms increases, the residuals become larger, such as seen in Fig. 4.8 and Fig. 4.10.

Table 4.2 shows a pronounced trend that the higher the incident WFE, the greater the reconstruction RMSE. This is clearly reflected using an example comparison between Fig. 4.5 and Fig. 4.6; the individual RMS WFE per modes increases in the each mode when the experiment was moved from low to high incident WFE. The same relationship is again confirmed in the correlation graphs for all 3 sets of Zernike terms with different incident RMS WFE, see in Fig. 4.4. Additionally, it is worth noting that increasing the number of Zernike modes used to construct the phase maps also leads to a higher reconstruction error within the same incident WFE regime. This is again reflected by the correlation graph, see Fig. 4.4, where there is a decrease in linear shape in the data clusters as the number of Zernike terms increases. As the number of Zernike terms in the phase mask increases, from 9 to 14 and finally 19, such as in Fig 4.6, Fig 4.8 and Fig 4.10, there is a visible increase in the reconstruction RMSE for each mode.

Fig. 4.12 shows the contribution to reconstruction RMSE (rad) from each mode, for each of the Zernike-basis experiments. While the distribution of error is roughly even, some trends can be discerned. In many cases (especially evident in the 9 and 19 Zernike plots) the first two modes, corresponding to tip and tilt, have higher errors. It is suspected this is due to drifts in alignment and vibration in the laboratory setup, introducing tip/tilt variation in the training data which is not reflected in training labels. The amount of the drift is quantified by comparing the change in total flux of the PL at its output using a series of equations below under the assumption that the initial system is perfectly aligned.

$$E_{x,y} = Ae^{-\frac{2x^2+2y^2}{\sigma^2}} \tag{4.1}$$

, where $I_{x,y}$ is the intensity at x, y position, A is the amplitude, x is the mean position at x axis, y is the mean position at y axis, σ is the standard deviation of Gaussian profile. The standard deviation can be replaced by the mode field diameter (the input beam diameter), w, by using the equation,

$$w = 2\sqrt{2\ln 2}\sigma \tag{4.2}$$

.

$$\eta = \frac{\int E_1 * E_2 dA}{\int E_1 dA + \int E_2 dA} \tag{4.3}$$

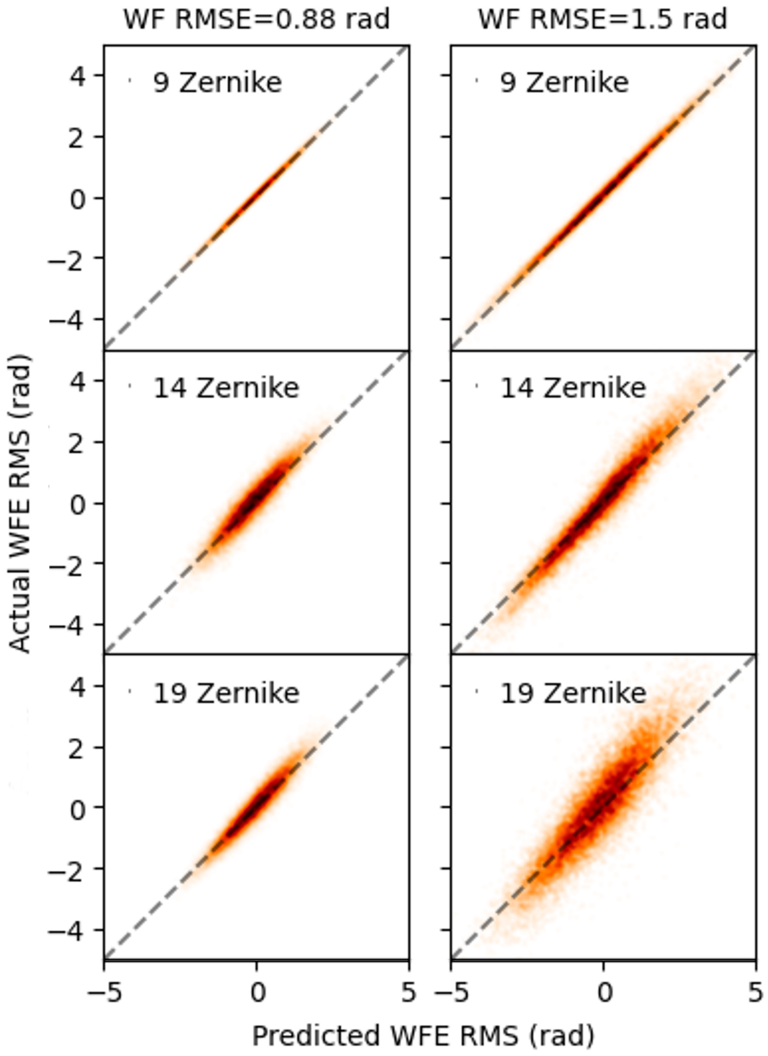


Figure 4.4: Correlation histograms between predicted WFE and actual WFE for the 6 sets of Zernike mode testing: the predicted coefficients closely match the actual coefficient values for 9, 14, 19 Zernike terms at lower incident WFE. As the number of Zernike term and RMS WFE increase, the scatter becomes larger. The dotted line is the theoretic best fit.

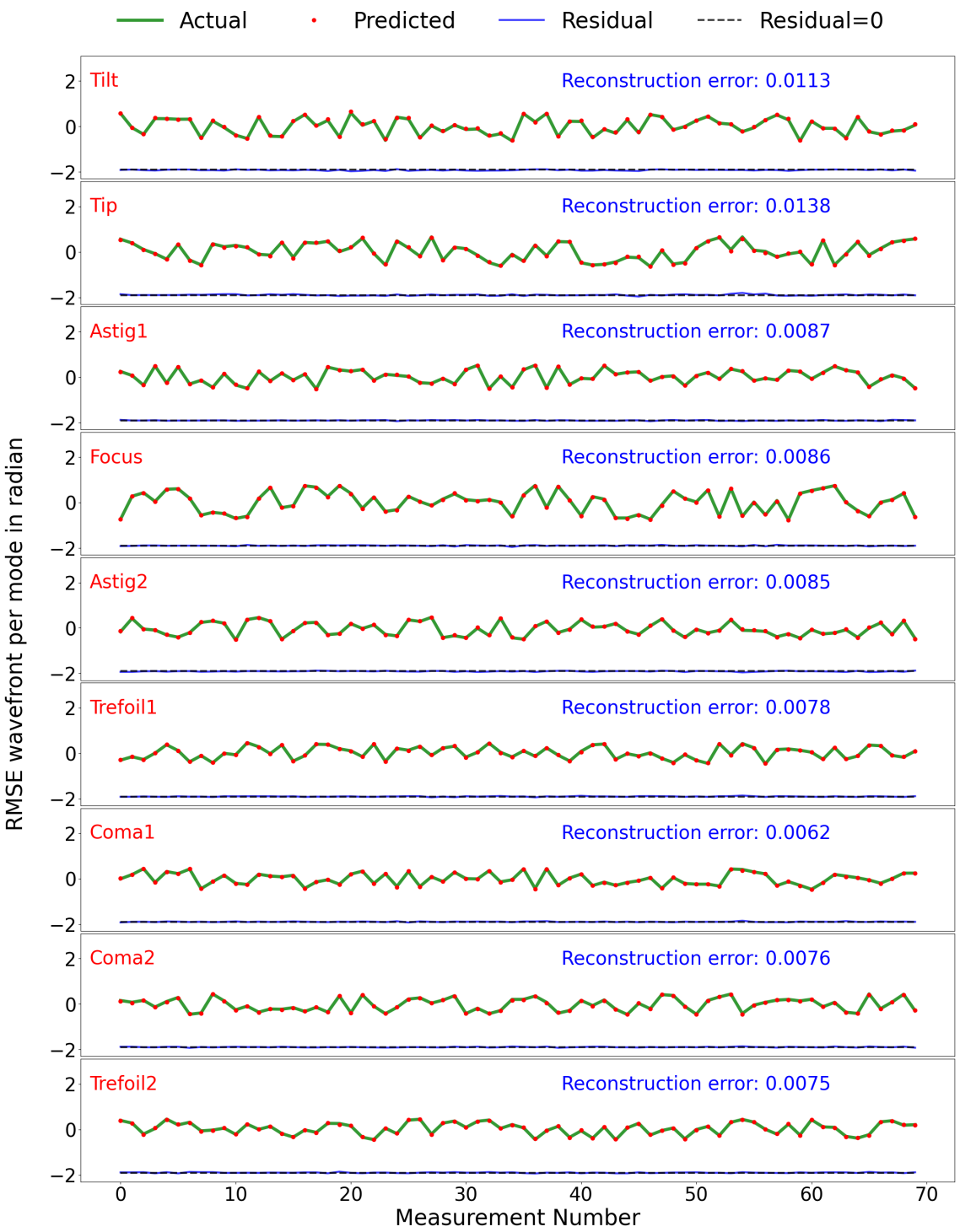


Figure 4.5: Results of wavefront prediction for the 9 Zernike phase masks with incident WFE of 0.88 rad RMS. Both true and predicted coefficients for all modes of 70 randomly selected data points are plotted. The corresponding residuals are also shown for each mode. The green line is the actual coefficient values, the red dot is the predicted values, the blue line is the residual and the dotted black line is the theoretic perfect fit. The residual between the actual and predicted values are almost overlapping with the perfect fit.

, where η is the coupling efficiency, E_1 is the PL mode field energy, E_2 is the beam mode field energy and dA is the common area between the two field. A function of η is established by integrating the dA over the mode area of the PL as seen in Fig. 4.11. The PL mode-field has a shape of a step function, therefore the convolution between the beam (a Gaussian profile) and the PL shape looks like an imperfect Guassian, shown in Fig. 4.11. The displacement in x, y direction is under the assumption that the output intensity from PL's MCF end has a Gaussian profile (happens when the intensity across the MMF end is assumed constant), however there is optical misalignment noted at the initial of each test case. The drift is calculated to be 0.03 micron/hour at best case and 0.12 micron/hour at the worst case across the 8 test cases. Also there appears to be a general trend where the higher order terms have lower errors. This may be because the effect of beam tip/tilt on the powered optics in the setup tends to induce lower order aberrations rather than higher order ones.

To allow a more visually intuitive presentation of the wavefront reconstruction, additional test sets were acquired where the Zernike coefficients varied continuously over time (produced by filtering noise by a Gaussian kernel), mimicking the continuously varying phase error experienced by a telescope in natural seeing, shown in Fig. 4.13 and Fig. 4.14.

It is worth noting that the performance of the NN is relatively poor for Tip and Tilt modes when compared to higher order modes. This could be attributed to the presence of background vibrations and thermal drifts in the laboratory setup, which result in modulation of the PSF position without a corresponding change in applied wavefront coefficients.

Two such sets of 1000 phase maps were generated with a incident RMS WFE of 0.88 and 1.5 rad respectively, and were injected into the PL and reconstructed using the existing 9-Zernike NN model. In Fig. 4.13 and Fig. 4.14, the PL and NN are seen to work well with the predicted coefficients coinciding with the actual coefficients for the 9 Zernike modes.

The same procedure was used to generate 2 time-series sets of 1000 phase masks with 19 Zernike terms (incident RMS WFE = 0.88 rad and incident RMS WFE = 1.5 rad respectively), were injected into the PL and reconstructed using the existing 19-Zernike NN model, see Fig. 4.15 and Fig. 4.16. It is as expected that the predictions for 19 Zernike terms are less accurate. Same prediction was shown for the previous data, seen in Fig. 4.9 and Fig. 4.10. Especially, the data with the high incident WFE (Fig. 4.16), shows the least accuracy in predicted coefficients per mode. There is a notable overall increase in reconstruction error per mode observed for both 9 and 19 Zernike terms in the two incident wavefront regimes comparing to their randomised counterparts in Fig. 4.5, Fig. 4.6, Fig. 4.9 and Fig. 4.10. This increase are likely attributed to the time difference (7 days) between the development of the model used for prediction and the acquisition of the time-series dataset. Over time, mechanical drift on the setup gradually introduced optical misalignment, leading to the observed increase in reconstruction error compared to the randomised 9 and 19 Zernike terms in both incident wavefront regimes.

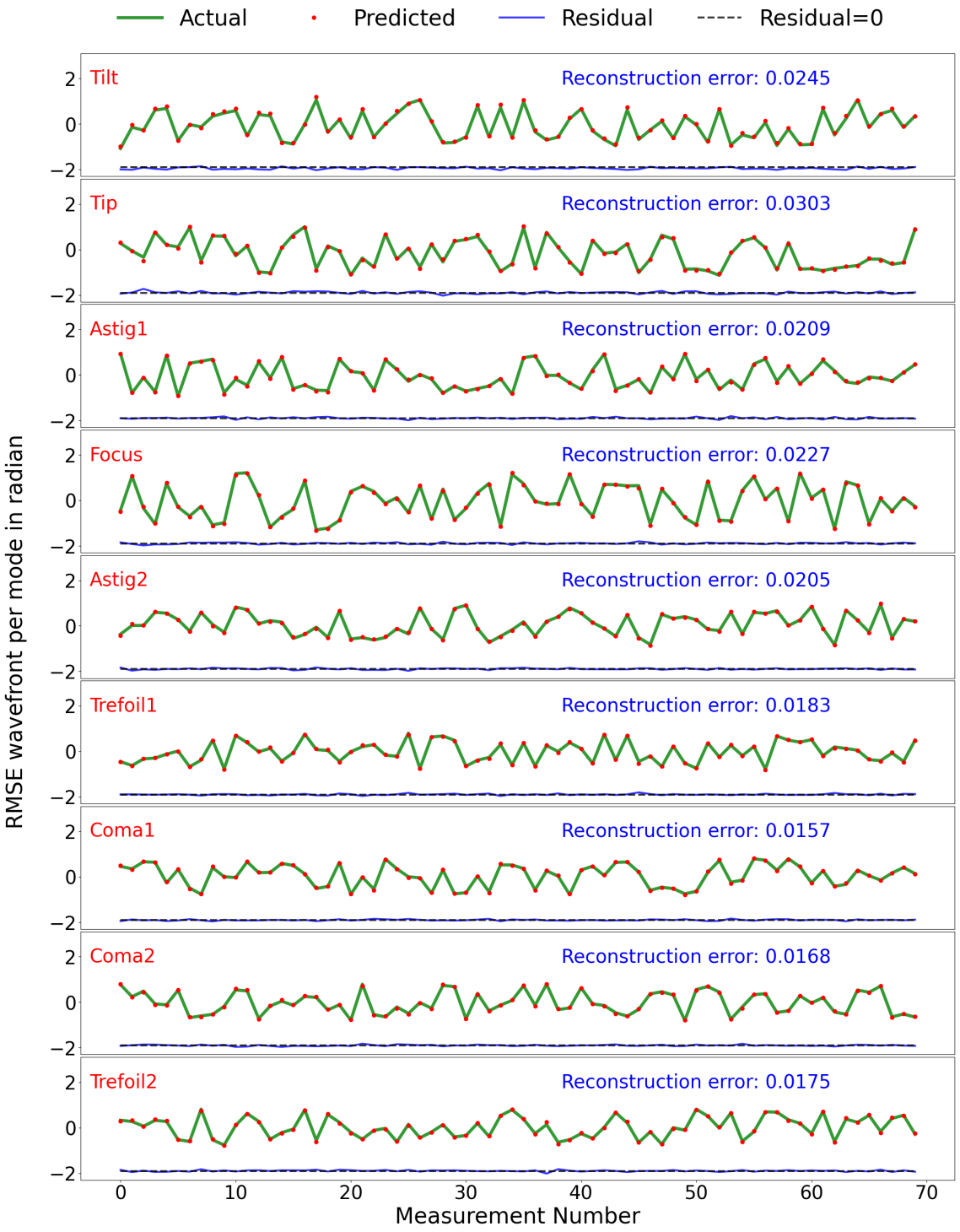


Figure 4.6: 9 Zernike phase masks with incident WFE of 1.5 in rad RMS: Both true and predicted coefficients for all modes of 70 randomly selected data points are plotted. The green line is the actual coefficient values, the red dot is the predicted values, the blue line is the residual and the dotted black line is the theoretic perfect fit. The residual between the actual and predicted values are still largely coinciding with the perfect fit, however there are noticeable departure comparing to the residual in Fig. 4.5.

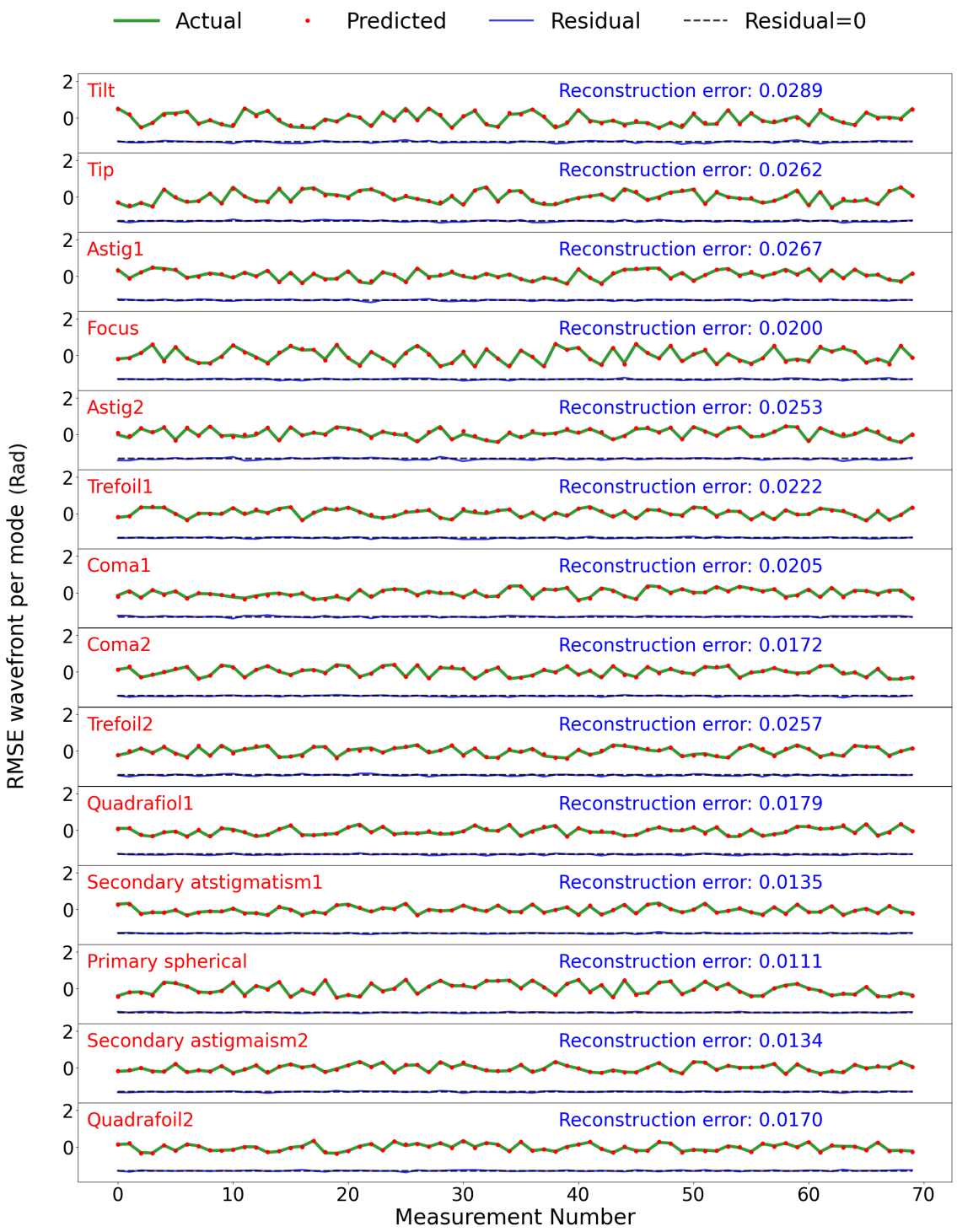


Figure 4.7: 14 Zernike phase map with 0.88 rad of incident RMS WFE: Both true and predicted coefficients for all modes of 70 randomly selected data points are plotted. The green line is the actual coefficient values, the red dot is the predicted values, the blue line is the residual and the dotted black line is the theoretic perfect fit. The residual between the actual and predicted values are largely coinciding with the perfect fit. There are noticeably disparities when comparing with Fig. 4.5, showing the reconstruction accuracy in this case is generally worse than the 9 Zernike term in the low WFE regime.

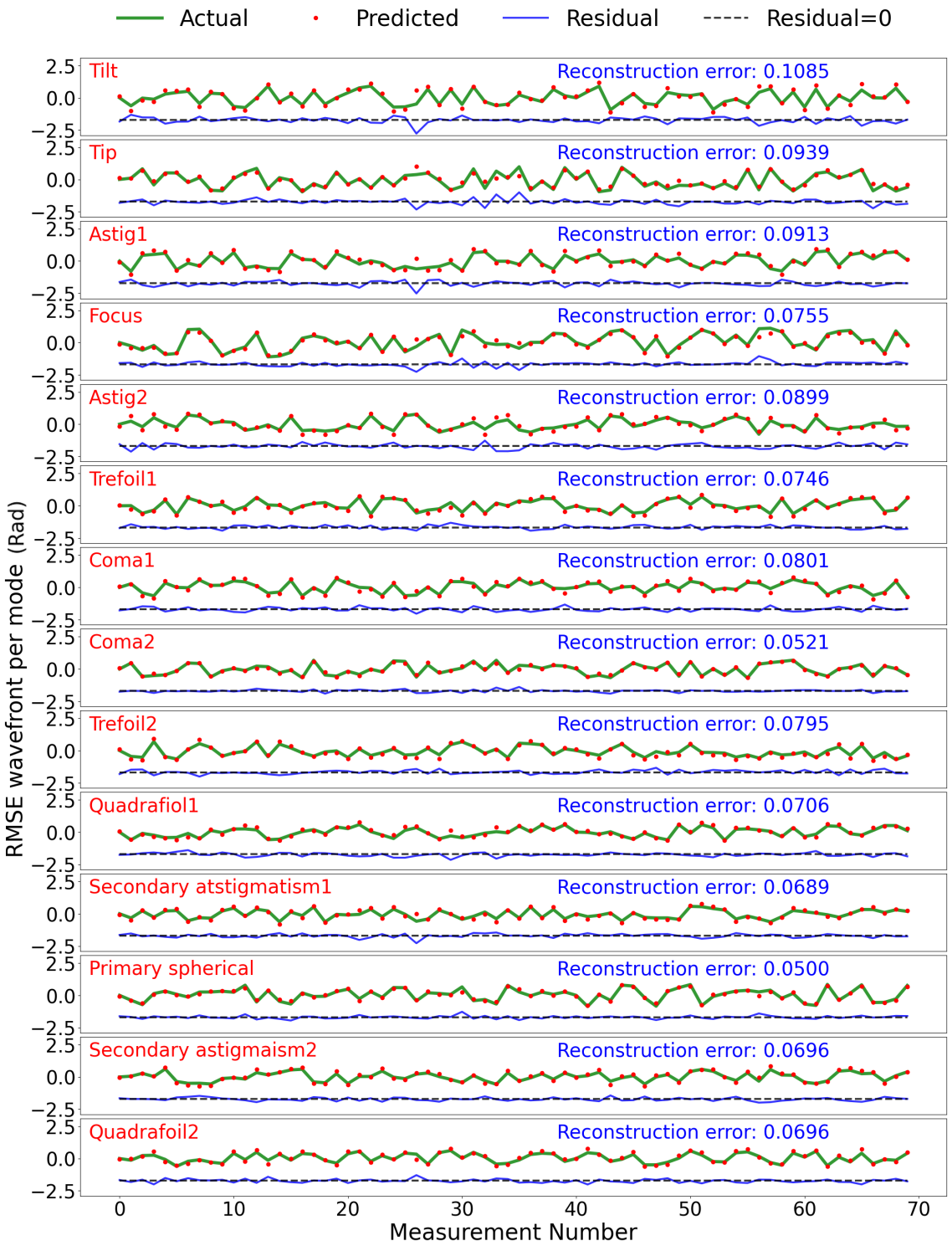


Figure 4.8: 14 Zernike phase map with 1.5 rad of incident RMS WFE: Both true and predicted coefficients for all modes of 70 randomly selected data points are plotted. The residual between the actual and predicted values are largely coinciding with the perfect fit. The disparities between residual and the perfect fit appears to be increasing comparing with Fig. 4.6 and 4.7, showing the reconstruction accuracy in this case is worse than the 14 Zernike term in the low WFE regime as well as the 9 Zernike terms in the high WFE regime.

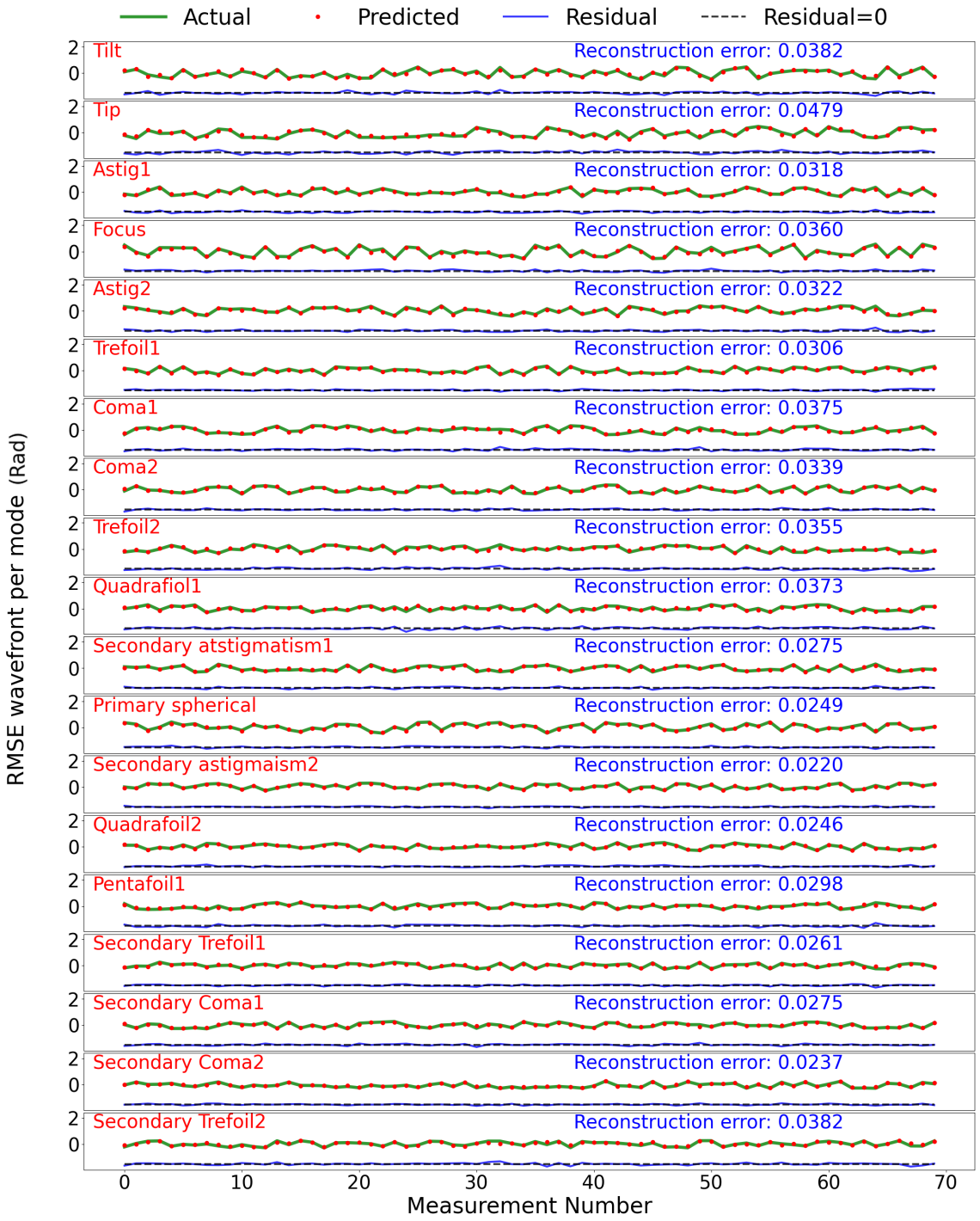


Figure 4.9: 19 Zernike phase map with 0.88 rad of incident RMS WFE: Both true and predicted coefficients for all modes of 70 randomly selected data points are plotted. The residual between the actual and predicted values are still coinciding with the perfect fit. The departure between residual and the perfect fit shows an further deterioration in reconstruction accuracy comparing with Fig. 4.5 and 4.7 in the same WFE regime. The disparity between residual and the perfect fit also more noticeable than those in Fig. 4.6, but comparable to those in Fig. 4.8, showing the reconstruction accuracy in this case is worse than the 9 and 14 Zernike terms in the low WFE regime as well as the 9 Zernike term in the high WFE regime.

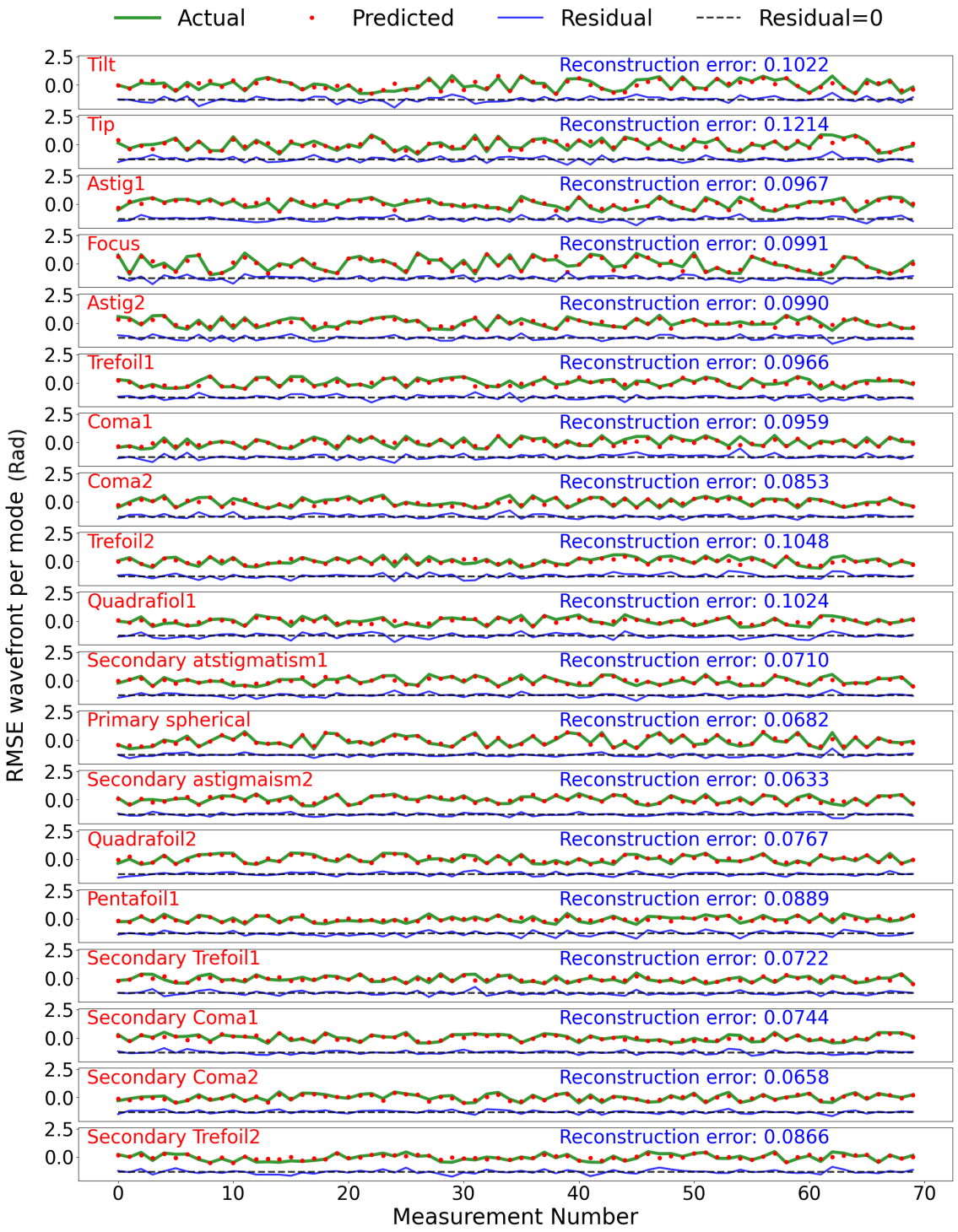


Figure 4.10: 19 Zernike phase map with 1.5 rad of incident RMS WFE: Both true and predicted coefficients for all modes of 70 randomly selected data points are plotted. The residual between the actual and predicted values are more departed from the perfect fit comparing with Fig. 4.6 and 4.8 in the same WFE regime. The disparity between residual and the perfect fit also is the most noticeable than those in the previous 5 cases. This shows the reconstruction accuracy in this case is worst in terms of wavefront reconstruction.

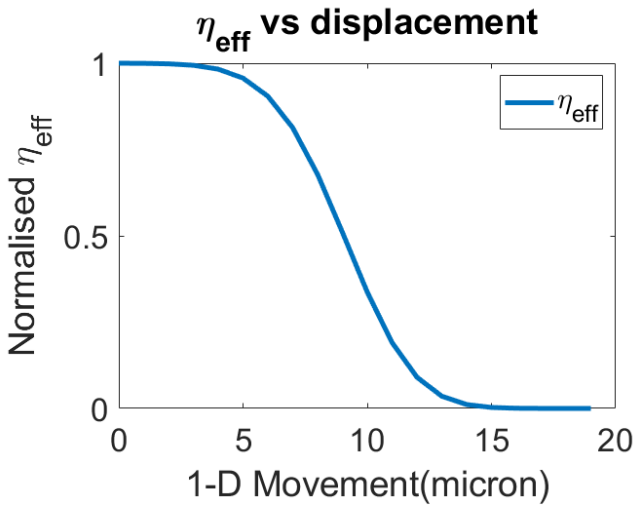


Figure 4.11: Function of coupling efficiency over change in common area between PL and input beam: as the beam move away from the centre of the PL mode field, an initial beam movement up to 5 micron corresponds to a very slow decline in coupling efficiency. A rapid drop in coupling efficiency is shown as beam moves over 5 micron. This coincides with the in the lab setting via change in the total output flux. The displacement of the optical system can be estimated via the change in output flux overtime.

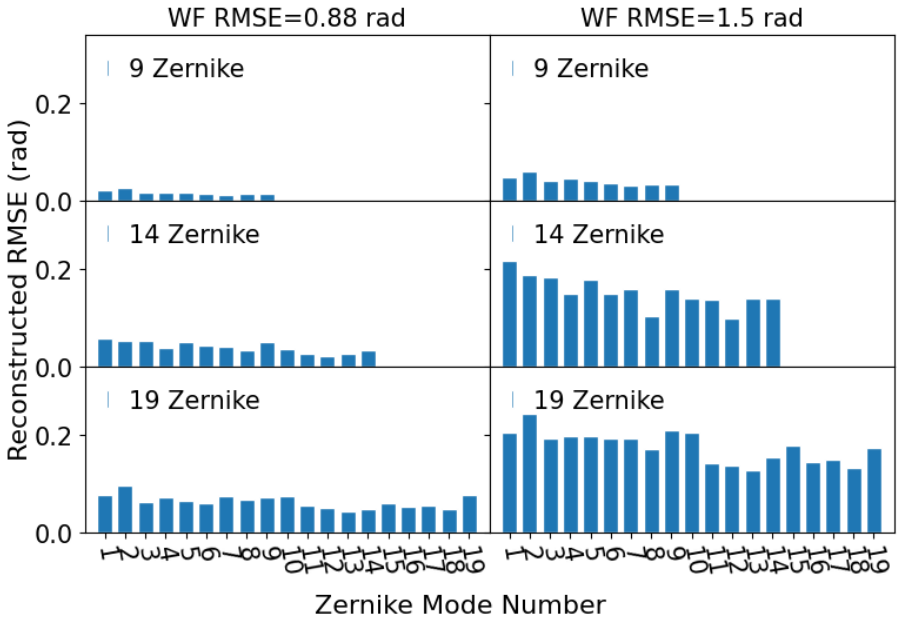


Figure 4.12: Reconstruction RMSE per mode for the Zernike-basis tests, for three different numbers of terms and for both the low and high incident WFE regime. There is a light decreasing trend in reconstruction error moving from lower order to higher order Zernikes. However due to the limited number of terms (19 terms), it is inclusive.

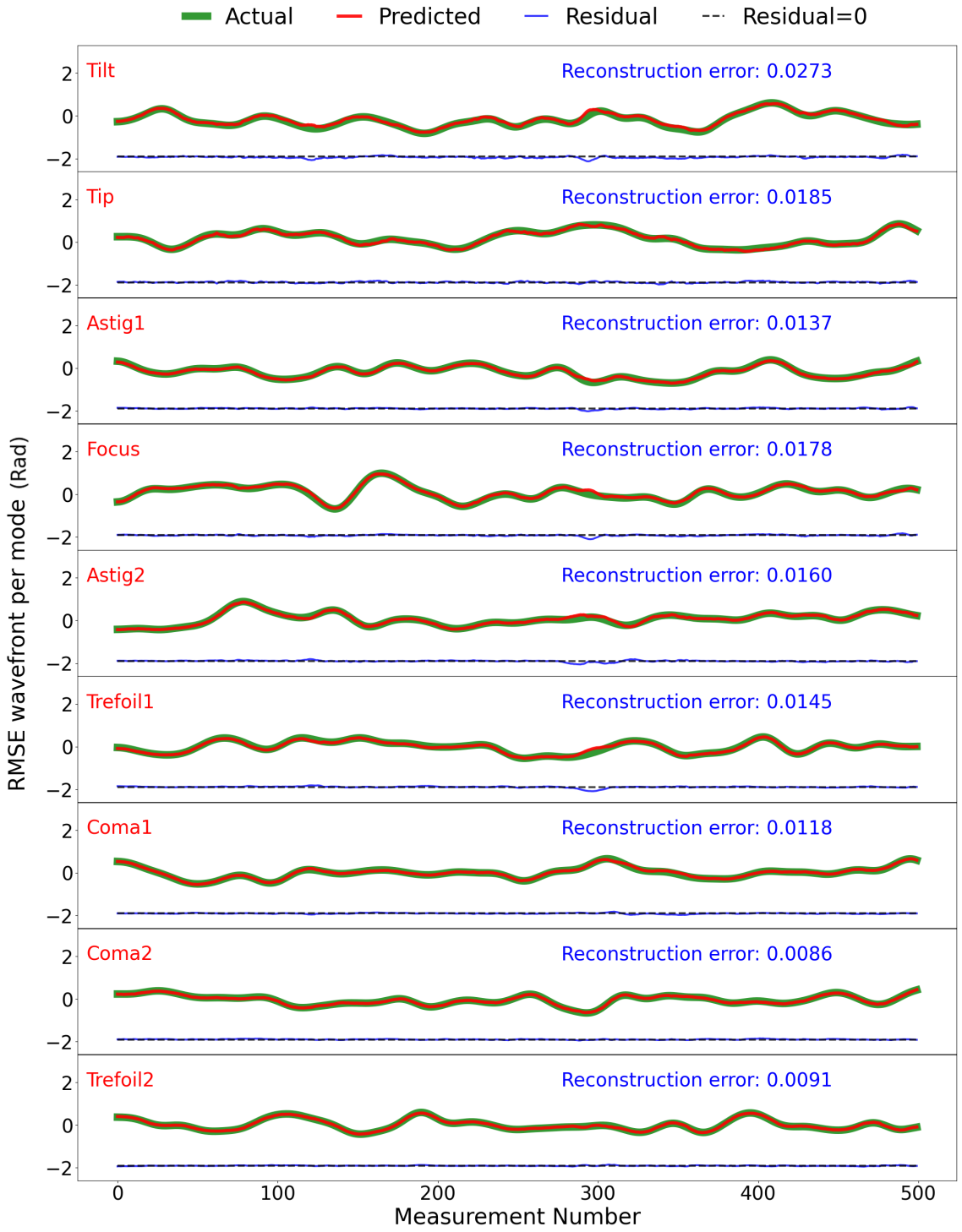


Figure 4.13: Results for a smoothly-varying time-series 9 Zernike phase map with 0.88 rad of incident RMS WFE: first 9 modes are shown (except piston mode) in a time series. The predicted coefficients almost coincide with the actual coefficient values in each for the 9 modes perfectly.

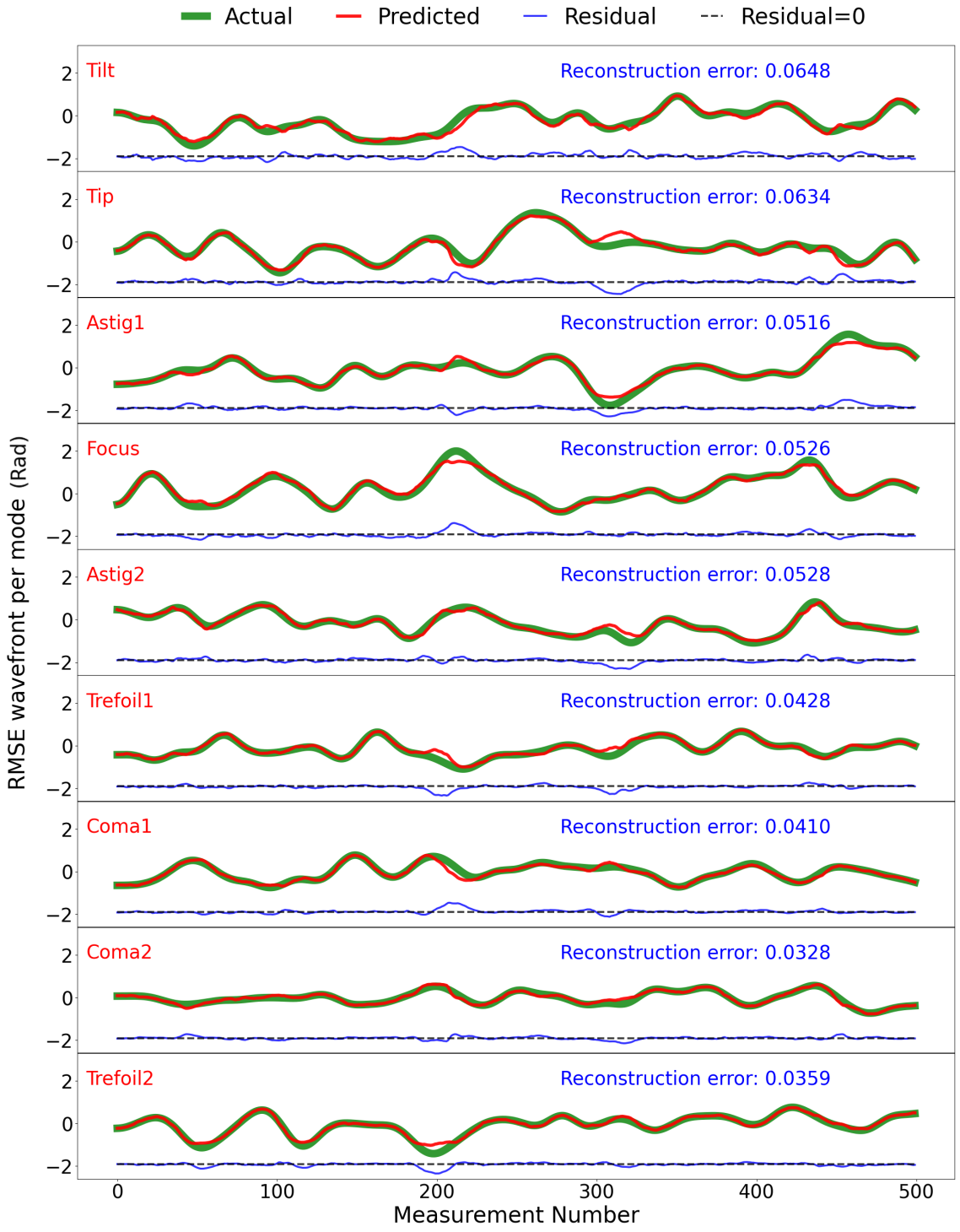


Figure 4.14: 9 Zernike phase map with 1.5 rad of incident RMS WFE in a time series: first 9 modes are shown (except piston mode) in a time series. The predicted coefficients still largely coincide with the actual coefficient values for each of the 9 modes.

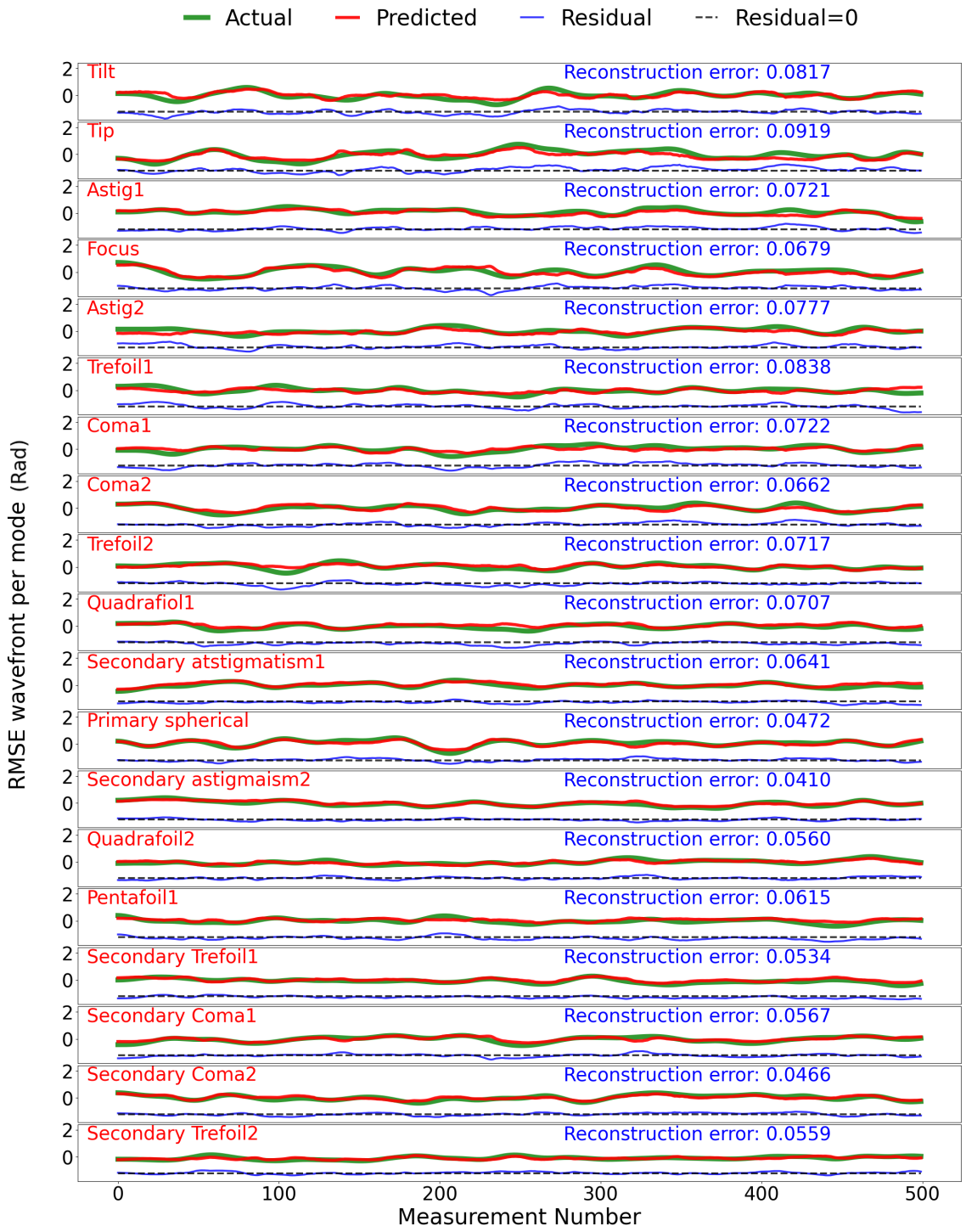


Figure 4.15: 19 Zernike phase map with 0.88 incident RMS WFE (rad) in a time series: first 19 modes are shown (except piston mode) in a time series. The predicted coefficients show deviation from the actual coefficient values in each of the 19 modes.

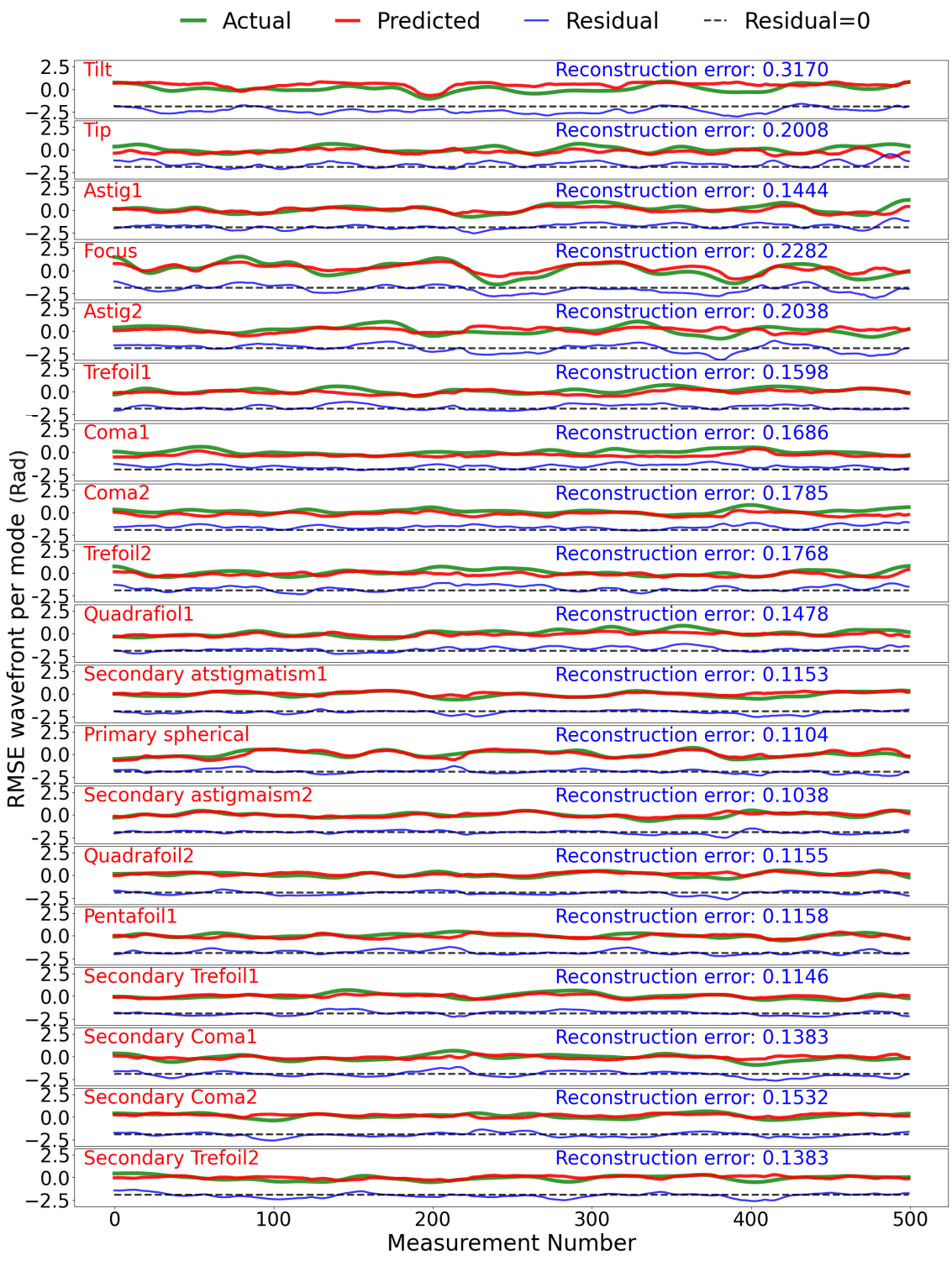


Figure 4.16: 19 Zernike phase map with 1.5 incident RMS WFE (rad) in a time series: first 19 modes are shown (except piston mode) in a time series. The predicted coefficients show further deviation with the actual coefficient values in each of the 19 modes.

4.3 Testing for Low Wind Effect

As mentioned previously, one of the advantages of a FP-WFS is the ability to detect the low wind effect (LWE) [18] also called the island effect (IE) [17]. The cause of this aberration was first confirmed by SPHERE using the Zernike sensor for Extremely Low-Level Differential Aberration(ZELDA)[17, 18].

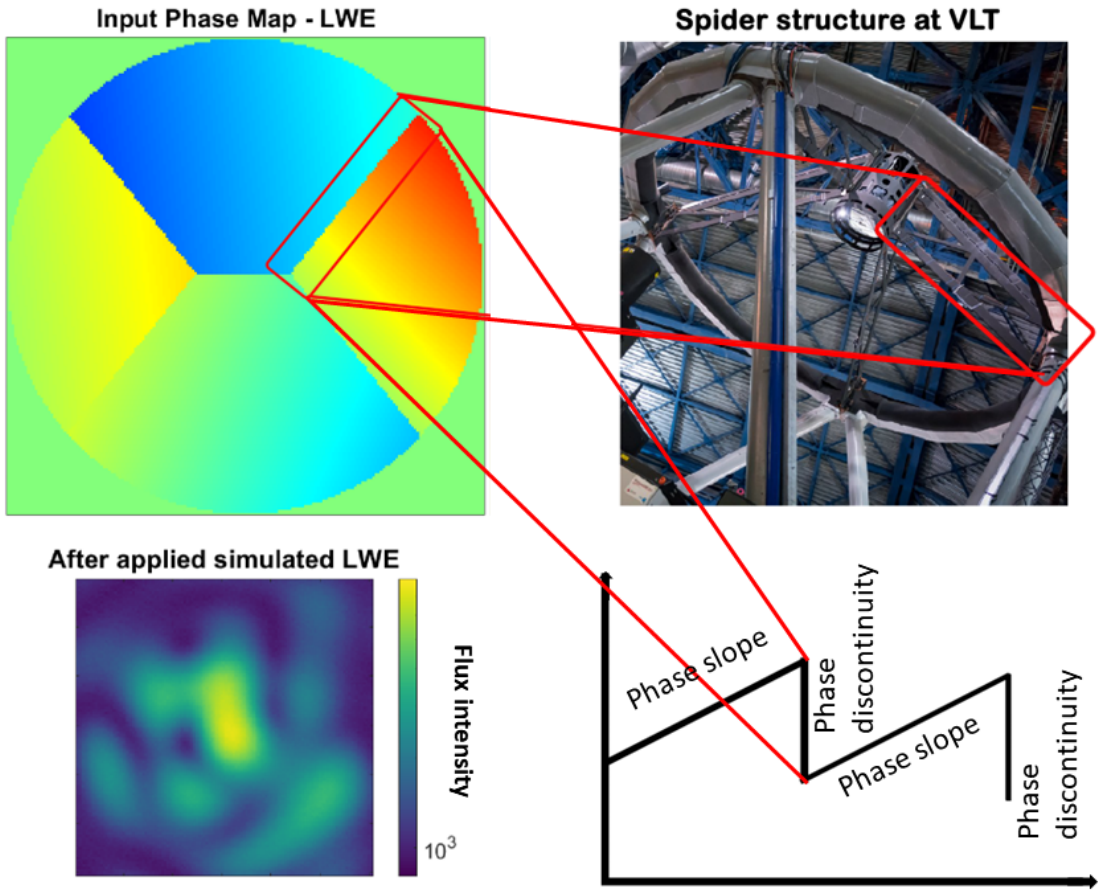
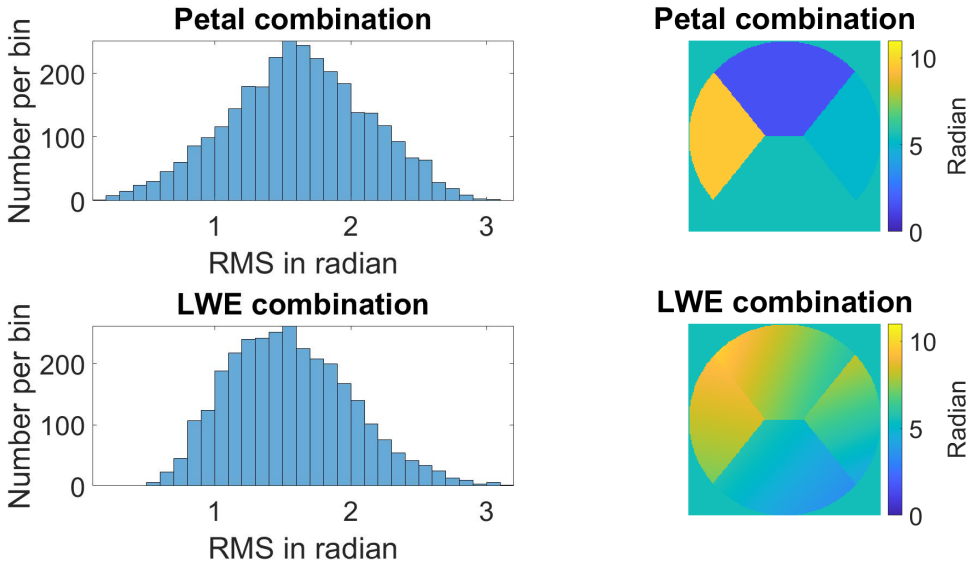


Figure 4.17: Phase slope and phase discontinuity: Top left shows simulated LWE on a phase map. Top right is an image of the Very Large Telescope(VLT) belonging to the European Southern Observatory (ESO) [72] located in Chile. Image credited to ESO/José Francisco (josefrancisco.org); Bottom left image captures what happens to the PSF of a point source when a LWE of 1.34 rad RMS WFE is applied to; Bottom right image shows the phase slope and phase discontinuity in a 2-D spatial domain.

Addressing this is also an opportunity to resolve another type of aberration mentioned previously, the NCP aberration. There are many methods to attempt to find solutions to these two aberrations, and efforts to resolve this issue are ongoing. [17]. Current PP-WFSs, such as a pyramid WFS, are limited in correcting this type of aberration due to the phase discontinuity at each spider beam [18] even in good seeing condition. 20% of the data were unusable due to the LWE on the SPHERE project [18]. This is a significant loss in cost-effectiveness given that the observing times are always limited and very competitive for the usage of any large telescopes



(a) incident WFE RMS

(b)a sample of phase map generated for each case

Figure 4.18: Distribution of total incident WFE in RMS (in rad) for petal and LWE phasemap: the top row is the petal WFE RMS distribution and a sample of the phase map; the bottom row is the LWE WFE RMS distribution and its sample phase map.

across different international projects. A FP-WFS, in theory, can detect this type of aberration well because the aberration has a very strong effect at the image plane [17, 25]. The first part of the experiment was to test whether the PL could observe phase discontinuity. A four-panel phase map was generated by applying piston a mode in each of the first 3 quadrants. The segmentation of the circle simulates the phase discontinuity in each spider beam Fig. 4.17.

The total incident WFE of each phase map was 1.5 rad RMS on average, seen in Fig. 4.18. 58578 sets of randomly generated phase maps were measured withe the PL, and the data were collected to feed into a NN with a 2 hidden layers (1000-100) with non-linear activation function to predict the coefficients of each of the piston modes. The calculation time for NN using GPU was 2.9 minutes in total. The predicted sets of coefficients of the same quantity were then compared with the actual sets of coefficients. The average reconstructed error per mode is 4.70×10^{-2} rad RMS, which corresponds to 9.1 nm in residual RMS WFE per mode. This demonstrates that the PL can see the phase discontinuity and predicts with high accuracy when a simple NN is also employed in data analysis.

For a more thorough invesitagion of LWE, the wavefront can be treated as a combination of piston, tip and tilt modes in each of the 4 quadrants[16, 17] in the simulation. The units of the coefficients are the gradients of tilts and tips, while the unit for piston modes is rad. Seen in Fig. 4.17, the circle is divided into 4 segments to simulate the 4 supporting beams of the spider structure over a pupil on a ground-based telescope shown in Fig. 4.17. The segmented pattern gives rise to phase discontinuity, which has been observed in large telescopes, caused by the

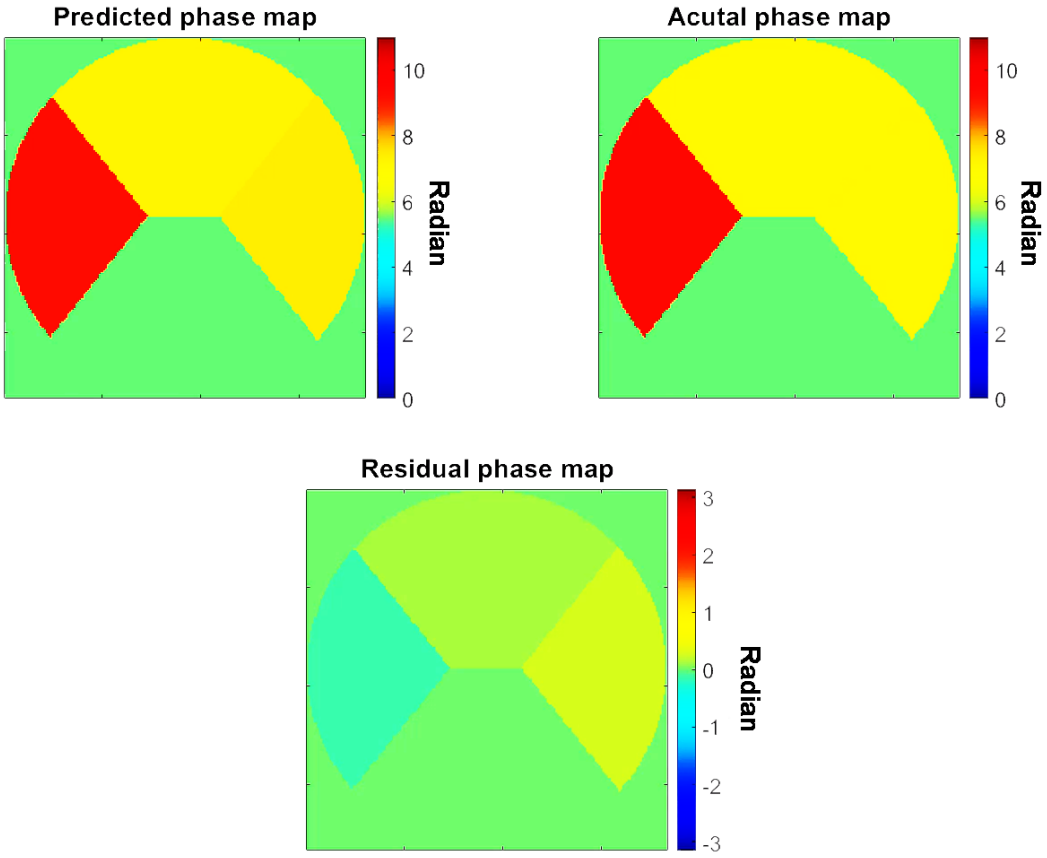


Figure 4.19: A comparison between the predicted phase map reconstructed by the predicted coefficients from NN and the actual phase map that PL had seen at the time of data collection : From left to right: top left shows a reconstructed phase map using the predicted result; the top right shows an actual phase map that PL saw during the experiment; The bottom panel shows a residual phase map subtracted between the two (note the different colour scale).

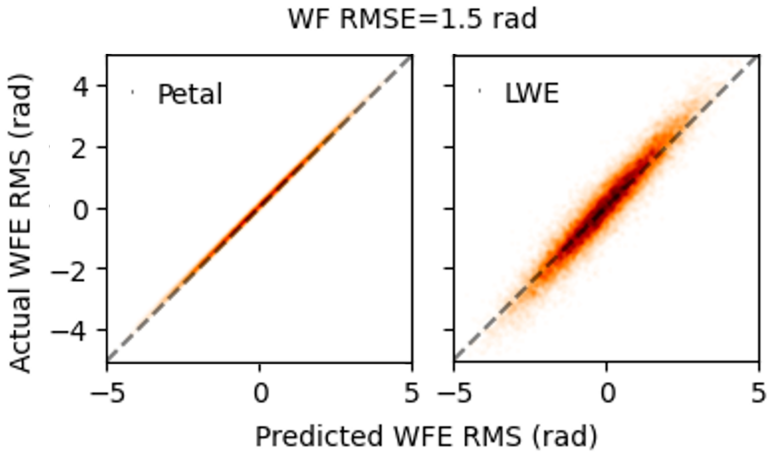


Figure 4.20: Correlation histogram between the actual WFE (in RMS rad) and the predicted WFE (in RMS rad) for petal and LWE mode testings: (on the left) the graph shows an almost linear correlation between the predicted data set using NN 3-layered structure and the actual total coefficients for the petal; (on the right) the correlation between the actual and predicted total coefficients of LWE. This graph appears to be less linear and more spread out, showing a less accuracy in reconstruction using an optimised NN due to the increased number of terms, N, in LWE phase masks (N=11) comparing to petal phase masks (N=3). The dotted line is the theoretic fit.

radiative effect of the spider beam (highlighted by the red box in the upper right image, Fig. 4.17). As mentioned earlier, PP-WFSs only detect the slope of the phase [18, 20, 26], this aberration cannot be well corrected by PP-WFS due to the shear of the phase (shown in the bottom right image in Fig. 4.17), although the aberration is strongly presented in the image plane (shown in the bottom left image in Fig. 4.17). The range of the randomly chosen coefficients for piston, tip and tilt modes was set such that the total WFE of each phase map is roughly centred at 1.5 rad RMS on average Fig. 4.18. The value of the RMS WFE was selected purely to test the PL's ability to detect wavefront in a large WFE regime. 3 of the 4 segments each contain a combination of the randomly generated tip, tilt, and piston modes to simulate the phase discontinuity caused by each supporting beam structure of the spider; the last panel left contains only tip and tilt, as it acts as the reference piston for the other 3 panels containing piston mode. Each individual phase mask , constructed by a total of 11 modes, is then displayed on the SLM to perform the PL measurement as seen in Fig. 4.22.

A set of 80000 randomly generated combinations of the 11 coefficients was used to construct the 8000 LWE phase maps, with each containing an average RMS WFE of 1.5 rad. The coefficients and the corresponding 19 fluxes of the PL MCF were used to train the NN. The NN used for predicting petal and LWE can be seen in Table 4.3). Both NN use 20% of the total data as validation, and were trained for the stated epochs shown in Table 4.3 using a total of 20 minutes. Evaluation with the validation data resulted in a reconstruction RMSE of 1.04×10^{-1} rad per mode, which is equivalent to a 11.6 nm residual RMS WFE per mode for the source wavelength at 700 nm. This results in a total RMS reconstruction error of 38.4 nm.

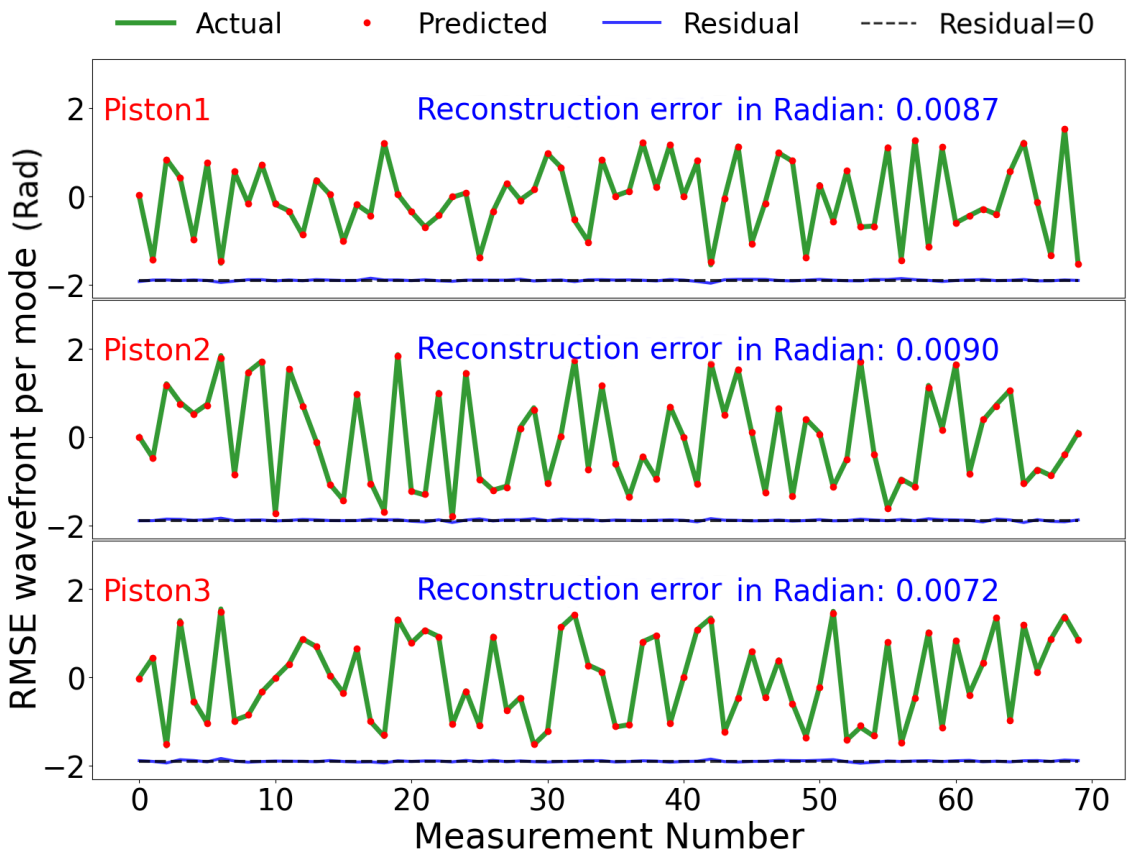


Figure 4.21: Comparison between the actual coefficients and predicted coefficients by the NN for petal modes with average input RMS WFE of 1.5 rad: the residuals of each modes are the blue line (almost indistinguishable from the axis).

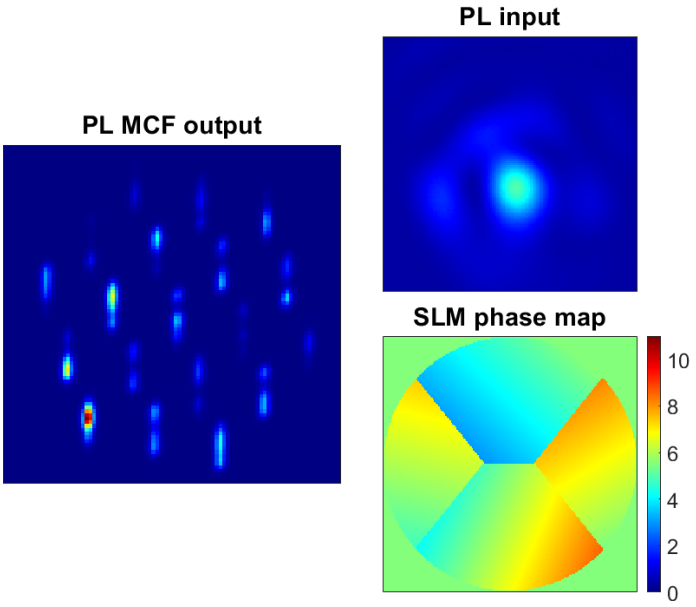


Figure 4.22: An example from the LWE test bed: The panel on the left is the image recorded from Cam1, seen in Fig 3.2. Note each of the 19 outputs consisting of an individual spectrum with 70 nm bandwidth (of which a smaller bandpass was selected for these experiments); the top right panel is the PSF after applying the phase map of the panel below.

Phasemap	NN layers with neuron number	Number of epoch	dropout rate	learning rate
Petal	1000-100-3	100	0.005	1×10^{-4}
LWE	12000-2000-2000-11	100	0.15	1×10^{-4}

Table 4.3: NN construct with hyperparameter for peddle and LWE with incident wavefront (RMSE=1.5 rad in average): all of the activation function for the NN is non-linear (ReLu)

The strong correlation between true and predicted coefficient values is shown in the left in Fig. 4.20.

Predicted WFEs that comprised the LWE phase map compared to the actual WFEs are shown on the right side of Fig. 4.20, while Fig. 4.23 shows the individual modes' comparisons in reconstruction RMSE (rad). Qualitatively, the differences between the predicted phase maps and the actual phase maps for petal and LWE respectively are almost visually negligible, seen in both Fig. 4.19 and Fig. 4.24. This is an excellent result, especially considering the very high incident WFE and high accuracy of the reconstruction compared to other work. For example, the asymmetric Fourier pupil wavefront sensor (APF-WFS) method mentioned in [17] only had an incident RMS WFE of < 1 rad in the lab-based test.

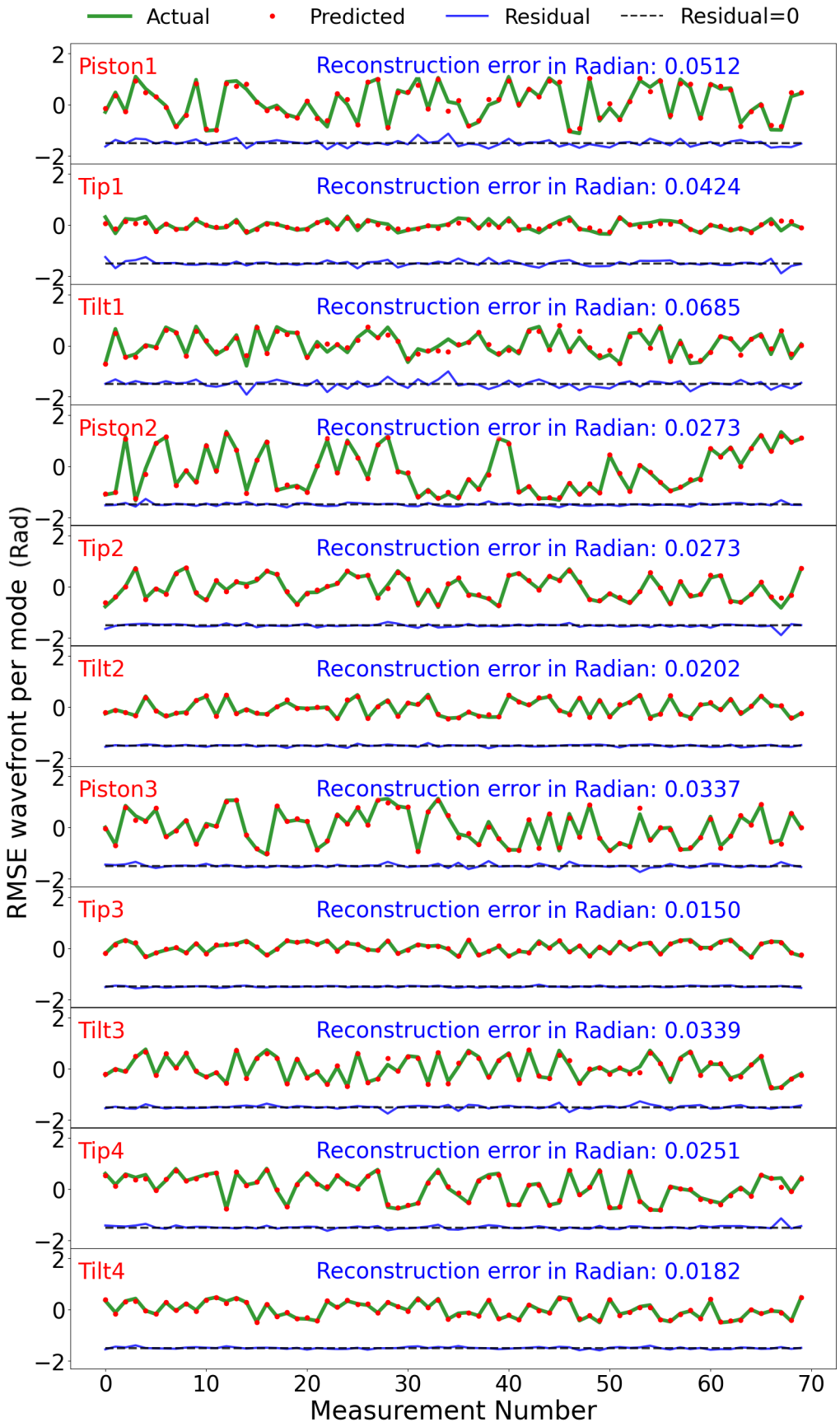


Figure 4.23: 70 data points for predicted and actual coefficients for the 11 individual mode that made up the LWE phase map with incident average WFE of 1.5 rad RMS. The total of residual RMS is displayed for each mode. The corresponding residuals are also shown for each mode.

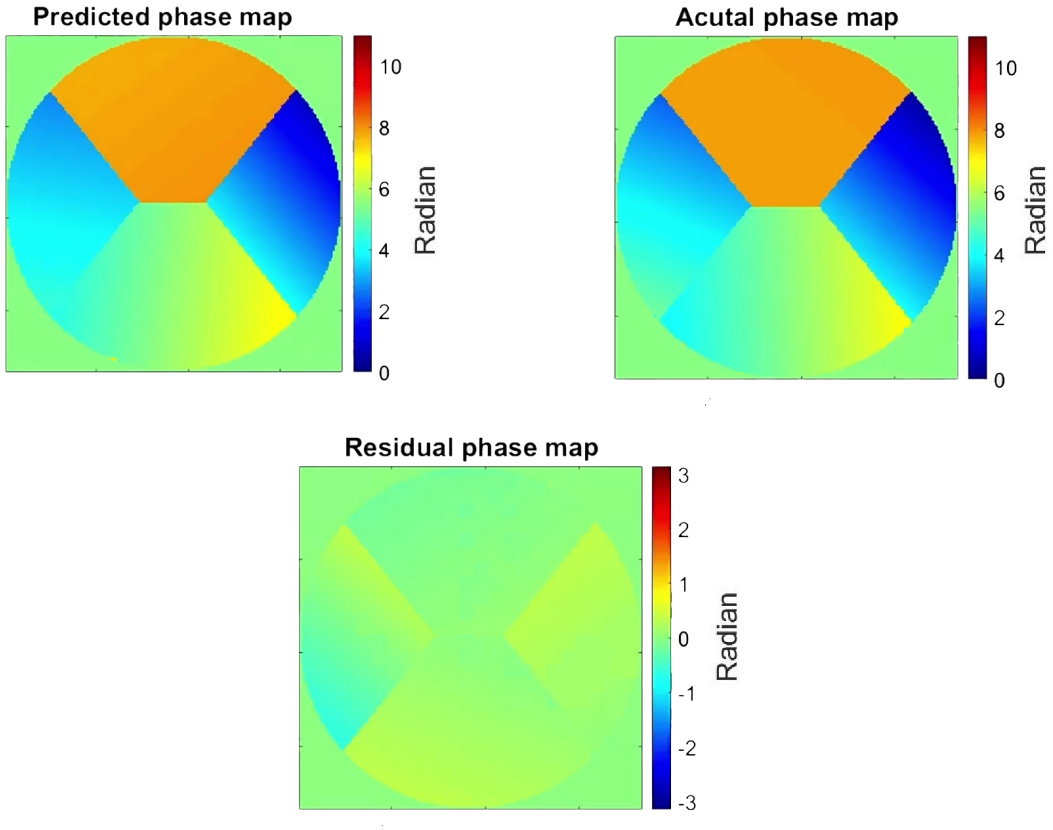


Figure 4.24: Comparison between the actual LWE phase map and predicted phase map using NN:Visual comparison between the reconstructed phase map using predicted coefficients and the original phase map. The panel at the bottom is the residual subtraction between the two.

4.4 Testing for Atmospheric Turbulence

The next step of the experiment is to use the atmospheric turbulence simulation generated by HCIPy [73] in the same setup as in the previous experiment. From the results in Section 4.2, the 9 Zernike phase map with low incident WFE gives the best result using NN. However, due to the complexity of a true atmospheric turbulence, 19 Zernike was selected as the basis in an attempt to fit the turbulence phase map. The low incident WFE regime is applied to both 19 Zernike phase map and turbulence phase map.

To minimise the impact of potential mechanical drifts and thermal effects from the surrounding environment on the optical setup, the input data training set was 74,280 phase masks of 19 Zernike phase maps, with sets of turbulence phase maps taken periodically. The average incident RMS WFE was set at 0.88 rad. The NN was trained in the same way as the previously described Zernike-basis networks using 80000 randomly generated Zernike phase maps (19 terms).

The simulated atmospheric phase screen contains higher spatial frequencies than the low orders sensed by the PL. To visually evaluate the PLs low-order reconstruction of the phase screen, a low-order representation of test phase screens were

reproduced by fitting the lowest 19 Zernike polynomials (using OpticsPy [74]), as shown in Fig.4.25. With the original phase screen applied to the SLM, the fluxes from the 19 output of the PL's MCF end were injected into the trained 19-Zernike NN model, which allowed the network to give the prediction in 19 Zernike terms according to the incident turbulence on SLM. The predicted coefficients were used to reconstruct the phase screen from these low-order Zernike terms. Very similar phase patterns can be seen in the predicted and the actual phase maps in Fig.4.25. This suggests that despite the high spatial frequency contained in the input SLM phase map, the PL produced a good prediction of the turbulence phase map using 19 Zernike modes as its basis in a laboratory demonstration, hence this method of detecting wavefront is feasible in an on-sky test in future development.

It is worth noting that despite the prediction of the PL using the NN shows close similarity to the true phase of the turbulence, there are larger variances at the edge of the pupil in the predictions of the PL shown in 4.25]. This may be due to the strong sensitivity of higher order Zernike modes to the very edge of the pupil, combined with any drift in alignment between the optical pupil location and the location of the pupil displayed on the SLM. Furthermore, quantitative comparisons of Fig. 4.25 (between the predicted and fitted Zernike phase maps) are not analysed due to the difference in the programs were employed, leading to discrepancies. The Zernike terms were created using the Matlab function - Zernfun, and the fitted Zernike terms were generated by applying the Python - Zernfit function which encoded with different sets of parameters. The objective of this comparison is solely qualitative in nature and never to test how well two different programs can model a known surface with Zernike polynomials as the basis. To sum up, the PL employed as a FP WFS demonstrates impressive accuracy in predicting higher frequency Kolmogorov phase maps when utilising a lower-order 19 Zernike modal bases. This further motivates future on-sky research into PL as a workable WFS.

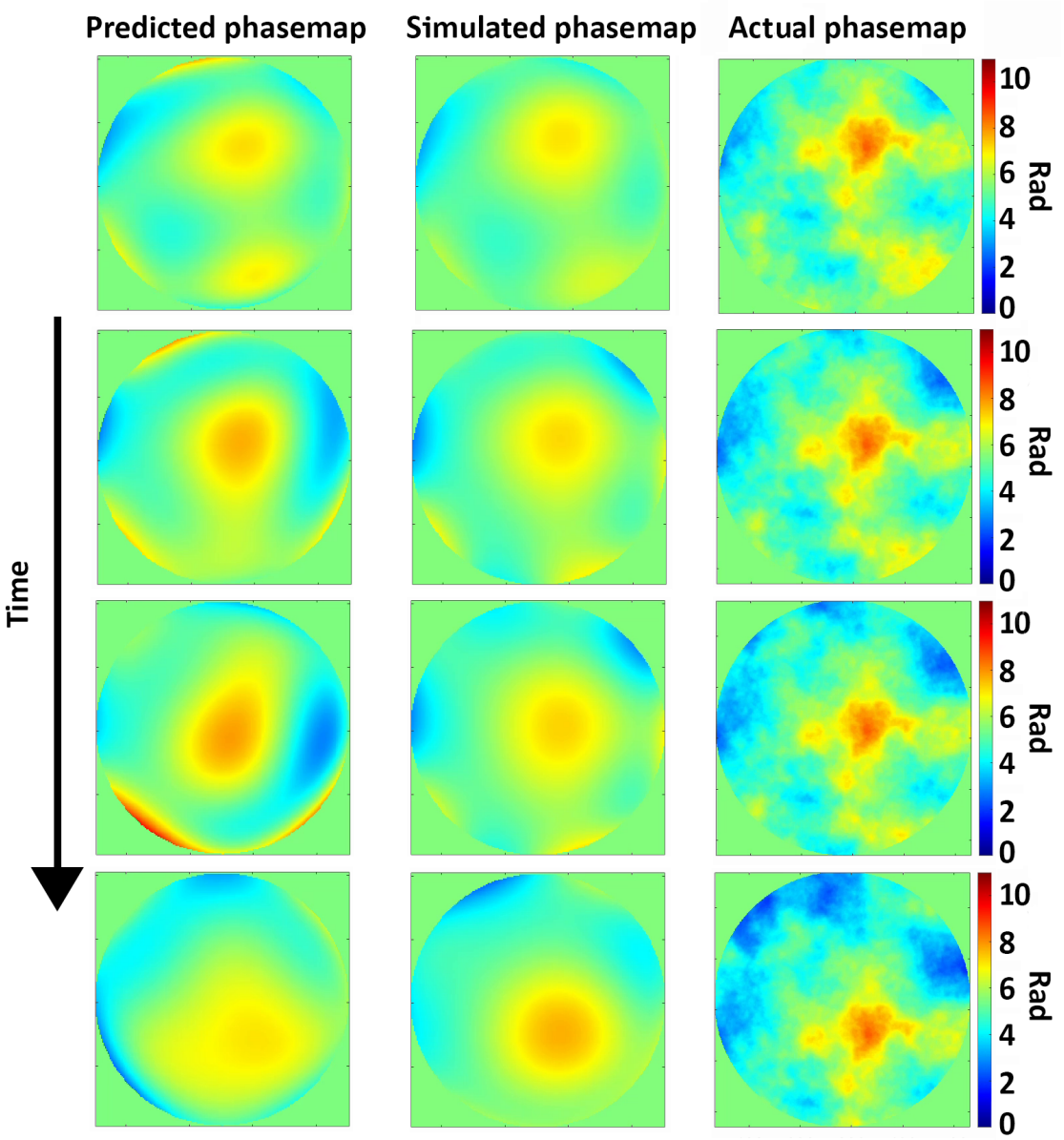


Figure 4.25: Comparison of a time series of the measurement of simulated atmospheric turbulence by the PL. Left: the prediction from the PL, centre: true phase map (fitted by low-order Zernike terms), and right: the true SLM phase mask of turbulence, which all display close similarity.

Chapter 5

Conclusions

WFSs play a crucial role in adaptive optics systems as they serve as the sensing tool to detect the incident phase of the incoming light source. The type of WFSs, PP-WFSs, used in contemporary AO systems are not sufficient by themselves to provide sufficient wavefront correction to resolve very high contrast objects such as rocky exoplanets next to their parent stars. Accurately detecting WFE and restoring the light source to its diffraction limit is crucial in detecting and analysing exoplanet formation and evolution, particularly for ground-based telescopes where atmospheric distortion caused by turbulence is constantly present. The limitations of currently working WFSs are partially due to their inability to detect errors such as non-common-path error as well as low-wind-effect and such limitation constrains the accuracy of the reconstructed phase values. The technical reason for such limitation has been discussed in Section 1.2 and 1.4. There have been numerous attempts to address these issues, either through the development of new WFSs or through post-digital processing techniques. However, these attempts have their own limitations and have yet to produce a working device for routine on-sky use. The concept of developing a new type of FP-WFS has been long desired in astronomy as this type possesses beneficial features complementary to PP-WFSs. These include eliminating NCP aberration and providing a large dynamic range for wave correction. The FP WFS is also desirable due to the potential to correct modes that are difficult for PP-WFSs to detect, such as LWE which is proven to be a serious problem for extreme AO systems in the era of very large telescopes. All these limitations and theory has been discussed in Section 1.4.

The work on this thesis proposes a new technology to tackle the limitations of current WFSs by focusing on a new type of WFS belonging to the FP-WFS category. The approach involves using a piece of existing photonic technology, a PL, that has been theorised for use as a WFS in recent years. It is thus advantageous as the fabrication of the PL has been refined and matured over the last several years with various studies characterising and improving the PL's different properties. The theory behind how PL can be used as a WFS has been discussed in Section 2.3. Furthermore, in this thesis we have shown that there is a very promising potential for PLs in conjunction with neural networks (NN) as functional FP-WFSs for future AO systems. The passive phase detection of a PL gives almost real-time sensing which minimises the potential of introducing more uncertainty into the system.

The previous works on other forms of FP-WFSs produced results using the linear approximation method to a small incident WFE ($\ll$ 1 radian RMS), which is less than can be experienced by ground-based telescopes. However, this thesis presents the result showing that the PL-WFS is able to detect incident RMS WFE (on average 1.5 rad; maximum at 3 rad), greater than many of its contemporary counterparts and with very small reconstruction RMSE. The PL-WFS is shown to be able to detect LWE, which is one of the developmental goals for new-generation WFSs for extreme AO as discussed in Section 1.4 and 1.3. We confirmed the success of the LWE detection for firstly with the test on petal modes for incident WFE RMS of 1.5 rad in the first part of Section 4.3. The reconstruction RMSE shows a very promising result as the total reconstruction RMSE of the 3 petal modes is 8.10×10^{-2} rad RMS. Furthermore, we also demonstrate the PL's ability to sense LWE at the same non-linear regime in the same section. The result is also significant as the total reconstruction RMSE in RMS is only 3.50×10^{-1} rad for a single iteration of correction at a very large incident WF in an experimental setting, which has not been shown in other published work for physical testing of any FP-WFS to the best of our knowledge.

The PL has further shown accurate wavefront sensing for up to the fourth-order Zernike phase mask (9 Zernike combinations) of the same incident WFE RMS in Section 4.2. This is already a significant result for the future implementation of this technology for wavefront sensing, since the majority of the aberrations happen and can be described in the lower order regime. The result is promising as it demonstrates a very small reconstruction RMSE and a relatively short processing time needed for a somehow simple NN on an off-the-shelf consumer computer. The limitation of the same quality wavefront sensing on higher-order terms, such are those in Section 4.2 is largely associated with the original number of the outputs of the PL used in the experiment. Despite the diminishing accuracy on predicting higher-order Zernike terms, the laboratory experiment confirms that a 19-output PL FP-WFS can predict up to 19 Zernike terms at a large incident WFE, while producing a total reconstruction RMSE between 1.85×10^{-1} rad to 9.10×10^{-1} rad for the different number of Zernike terms (9, 14 and 19), in a phase mask. It is therefore anticipated that this will scale well as PLs with more outputs are used.

Furthermore, we have also demonstrated the broadband nature of this technology. The light source used for the thesis was broadband as shown in Section 3.2. Broadband operation is associated with higher input intensity, lower speckle at the image plane, as well as (when wavelengths are separately resolved) resolving phase wrapping problems in monochromatic sources. Early results shows that broadband operation and wavelength diversity will offer an unique advantage for future developments of this technology as a wavefront sensor.

This thesis is only the starting point of the study of PLs as WFSs. There is more work needed to fully exploit and characterise this technology, as well as to develop full deployable WFSs for future generation AO systems. Many directions can be taken to further the study this technology. For example, using a PL as FP-WFS also provides further convenient integration into other astronomical systems such as exoplanet spectrographs, sky-glow suppression instruments, and astronomical interferometers to name a few. For instance, the direct integration of diffraction-limited

spectroscopy and a FP-WFS may provide a more compact instrument that provides a much more powerful instrument for analysing exoplanet atmosphere composition in order to search for biological signatures. The PL as FP-WFS can also function as a novel photonic direct imaging sensor that could both detect phase and intensity simultaneously while removing NCP aberration - a significant problem in current imaging AO systems.

The limits of this technology with respect to the number of modes for a PL to increase detection of higher order aberrations have yet to be investigated. Further studies in both experimental and theoretical efforts towards a full understanding of PL's capability as FP-WFS is an exciting prospect and no doubt highly beneficial for optical sensing and imaging technology as a whole. The data acquisition and AI components of the PL WFSs, such as the NN framework and deployment, will also need to be developed and optimised for the processing time when integrated into working real-time AO systems. For instance, the integration of the NN as processing software is yet to be explored in an on-sky survey. Ultimately, we are just scratching the surface of the potential of this technology that will no doubt find important applications in wavefront sensing and optical imaging.

Bibliography

[1] Momen Diab and Stefano Minardi. On the modal throughput of photonic lanterns in the presence of partial adaptive optic correction. In Ramón Navarro and Roland Geyl, editors, *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III*, volume 10706, pages 1370 – 1376. International Society for Optics and Photonics, SPIE, 2018. doi: 10.1117/12.2309657. URL https://doi.org/10.1117/12.2309657.

[2] R.K. Tyson. *Introduction to Adaptive Optics*. Online access with subscription: SPIE Digital Library. Society of Photo Optical, 2000. ISBN 9780819435118. URL https://books.google.com.au/books?id=qbMZOkkNmuEC.

[3] Paul Hickson. Atmospheric and adaptive optics. *The Astronomy and Astrophysics Review*, 22:1–38, 11 2014. doi: 10.1007/s00159-014-0076-9.

[4] F. Roddier. Interferometric imaging in optical astronomy. Phys. Rep., 170: 97–166, November 1988. doi: 10.1016/0370-1573(88)90045-2.

[5] Richard Davies and Markus Kasper. Adaptive Optics for Astronomy. *Annual Review of Astronomy and Astrophysics*, 50(1):305–351, September 2012. doi: 10.1146/annurev-astro-081811-125447. URL http://www.annualreviews.org/doi/10.1146/annurev-astro-081811-125447.

[6] R.K. Tyson and B.W. Frazier. *Field Guide to Adaptive Optics*. SPIE field guides. Society of Photo Optical, 2012. ISBN 9780819490179. URL https://books.google.com.au/books?id=_otytgAACAAJ.

[7] Jean-François Sauvage, Thierry Fusco, Gérard Rousset, and Cyril Petit. Calibration and precompensation of noncommon path aberrations for extreme adaptive optics. *Journal of the Optical Society of America A*, 24(8):2334–2346, Aug 2007. doi: 10.1364/JOSAA.24.002334.

[8] Frantz Martinache. The Asymmetric Pupil Fourier Wavefront Sensor. PASP, 125(926):422, Apr 2013. doi: 10.1086/670670.

[9] Timothy R. Corle and Gordon S. Kino. Chapter 1 - introduction. In Timothy R. Corle and Gordon S. Kino, editors, *Confocal Scanning Optical Microscopy and Related Imaging Systems*, pages 1–66. Academic Press, Burlington, 1996. ISBN 978-0-12-408750-7. doi: https://doi.org/10.1016/B978-012408750-7/50009-4. URL https://www.sciencedirect.com/science/article/pii/B9780124087507500094.

[10] Julian C. Christou, Kenneth J. Mighell, and Russell B. Makidon. Strehl ratio and image sharpness for adaptive optics. In Brent L. Ellerbroek and Domenico Bonaccini Calia, editors, *Advances in Adaptive Optics II*, volume 6272, pages 628 – 639. International Society for Optics and Photonics, SPIE, 2006. doi: 10.1117/12.670430. URL https://doi.org/10.1117/12.670430.

[11] Alexander Komech and Anatoli Merzon. *Fresnel–Kirchhoff Diffraction Theory*, pages 19–35. Springer International Publishing, Cham, 2019. ISBN 978-3-030-26699-8. doi: 10.1007/978-3-030-26699-8_3. URL https://doi.org/10.1007/978-3-030-26699-8_3.

[12] Introduction to the electromagnetic spectrum, 2023. URL https://science.nasa.gov/ems/01_intro.

[13] Olivier Guyon. Extreme adaptive optics. *Annual Review of Astronomy and Astrophysics*, 56(1):315–355, 2018. doi: 10.1146/annurev-astro-081817-052000. URL https://doi.org/10.1146/annurev-astro-081817-052000.

[14] Olivier Lardiere, Marcel Carbillet, Armando Riccardi, and Piero Salinari. High-contrast imaging with ELTs: effects of cophasing and AO residual errors on the PSF contrast. In Domenico Bonaccini Calia, Brent L. Ellerbroek, and Roberto Ragazzoni, editors, *Advancements in Adaptive Optics*, volume 5490, pages 516 – 526. International Society for Optics and Photonics, SPIE, 2004. URL https://doi.org/10.1117/12.550403.

[15] Jacques M. Beckers. Adaptive optics for astronomy: Principles, performance, and applications. *Annual Review of Astronomy and Astrophysics*, 31(1):13–62, 1993. doi: 10.1146/annurev.aa.31.090193.000305. URL https://doi.org/10.1146/annurev.aa.31.090193.000305.

[16] J. Milli, M. Kasper, P. Bourget, C. Pannetier, D. Mouillet, J. F. Sauvage, C. Reyes, T. Fusco, F. Cantalloube, K. Tristam, Z. Wahhaj, J. L. Beuzit, J. H. Girard, D. Mawet, A. Telle, A. Vigan, and M. N'Diaye. Low wind effect on VLT/SPHERE: impact, mitigation strategy, and results. In Proc. SPIE, volume 10703 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 107032A, Jul 2018. doi: 10.1117/12.2311499.

[17] M N'Diaye, F Martinache, N Jovanovic, J Lozi, O Guyon, B Norris, A Ceau, and D Mary. Calibration of the island effect: Experimental validation of closed-loop focal plane wavefront control on Subaru/SCExAO. *Astronomy and Astrophysics*, 610:A18, February 2018. doi: 10.1051/0004-6361/201731985. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2018A%26A...610A..18N&link_type=EJOURNAL.

[18] Jean-François Sauvage, Thierry Fusco, Masen Lamb, Julien Girard, Martin Brinkmann, Andres Guesalaga, Peter Wizinowich, Jared O'Neal, Mamadou N'Diaye, Arthur Vigan, David Mouillet, Jean-Luc Beuzit, Markus Kasper, Miska Le Louarn, Julien Milli, Kjetil Dohlen, Benoît Neichel, Pierre Bourget, Pierre Haguenauer, and Dimitri Mawet. Tackling down the low wind effect on

SPHERE instrument. In Enrico Marchetti, Laird M. Close, and Jean-Pierre Véran, editors, *Adaptive Optics Systems V*, volume 9909, page 990916. International Society for Optics and Photonics, SPIE, 2016. doi: 10.1117/12.2232459. URL https://doi.org/10.1117/12.2232459.

[19] Masen Lamb, David Andersen, Jean-Pierre Véran, Carlos Correia, Glen Herriot, Matthias Rosensteiner, and Jason Fiege. Non-common path aberration corrections for current and future ao systems. *Proceedings of SPIE - The International Society for Optical Engineering*, 9148:914857, 07 2014. doi: 10.1117/12.2055198.

[20] Barnaby Norris, Jin Wei, Christopher Betters, Alison Wong, and Sergio Leon-Saval. An all-photonic focal-plane wavefront sensor. *Nature Communications*, 11:5335, 03 2020. doi: 10.1038/s41467-020-19117-w. URL https://doi.org/10.1038/s41467-020-19117-w.

[21] Ben C Platt and Roland Shack. History and principles of shack-hartmann wavefront sensing. *Journal of refractive surgery*, 17(5):S573–S577, 2001.

[22] Roberto Ragazzoni. Pupil plane wavefront sensing with an oscillating prism. *Journal of Modern Optics*, 43(2):289–293, Feb 1996. doi: 10.1080/09500349608232742.

[23] Kevin Gurney. *An Introduction to Neural Networks*. Taylor & Francis, Inc., USA, 1997. ISBN 1857286731.

[24] Sebastien Vievard, Steven Bos, Frederic Cassaing, Alban Ceau, Olivier Guyon, Nemanja Jovanovic, Christoph U. Keller, Julien Lozi, Frantz Martinache, Aurelie Montmerle-Bonnefois, Laurent Mugnier, Mamadou NDiaye, Barnaby Norris, Ananya Sahoo, Jean-Francois Sauvage, Frans Snik, Michael J. Wilby, and Alisson Wong. Overview of focal plane wavefront sensors to correct for the low wind effect on subaru/scexao, 2019. URL https://arxiv.org/abs/1912.10179.

[25] Barnaby R M Norris, Nick Cvetojevic, Tiphaine Lagadec, Nemanja Jovanovic, Simon Gross, Alexander Arriola, Thomas Gretzinger, Marc-Antoine Martinod, Olivier Guyon, Julien Lozi, and et al. First on-sky demonstration of an integrated-photonic nulling interferometer: the glint instrument. *Monthly Notices of the Royal Astronomical Society*, 491(3):4180–4193, Nov 2019. ISSN 1365-2966. doi: 10.1093/mnras/stz3277. URL http://dx.doi.org/10.1093/mnras/stz3277.

[26] Olivier Guyon. High Sensitivity Wavefront Sensing with a Nonlinear Curvature Wavefront Sensor. PASP, 122(887):49, January 2010. doi: 10.1086/649646.

[27] Joseph W Goodman. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, 1, 2005.

[28] Xinxue Ma and Jianli Wang. The research of wavefront sensor based on focal plane and pupil plane. *Optik*, 127(5):2688–2693, 2016. ISSN 0030-4026. doi: https://doi.org/10.1016/j.ijleo.2015.10.141. URL https://www.sciencedirect.com/science/article/pii/S0030402615015090.

[29] Ji Wang, Dimitri Mawet, Jonathan J. Fortney, Callie Hood, Caroline V. Morley, and Björn Benneke. Detecting Water in the Atmosphere of HR 8799 c with L-band High-dispersion Spectroscopy Aided by Adaptive Optics. AJ, 156(6): 272, Dec 2018. doi: 10.3847/1538-3881/aae47b.

[30] R. A. Gonsalves. Phase Retrieval And Diversity In Adaptive Optics. *Optical Engineering*, 21(5):829, Oct 1982. doi: 10.1117/12.7972989.

[31] Visa Korkiakoski, Christoph U Keller, Niek Doelman, Matthew Kenworthy, Gilles Otten, and Michel Verhaegen. Fast & Furious focal-plane wavefront sensing. *Applied Optics*, 53(20):4565, 2014. doi: 10.1364/AO.53.004565. URL https://www.osapublishing.org/abstract.cfm?URI=ao-53-20-4565.

[32] R.S. Quimby. *Photonics and Lasers: An Introduction*. Wiley, 2006. ISBN 9780471791584. URL https://books.google.com.au/books?id=82f-gIvtC7wC.

[33] Bahaa E. A. Saleh and Malvin Carl Teich. *Fundamentals of photonics: Part 1: Optics; 3rd edition*. Wiley series in pure and applied optics. Wiley, Hoboken, 2019. ISBN 9781119506867. URL https://bib-pubdb1.desy.de/record/407725.

[34] J. Tyndall. *Six Lectures on Light: Delivered in America in 1872-1878*. D. Appleton, 1886. URL https://books.google.com.au/books?id=PRkxwQEACAAJ.

[35] Joss Bland-Hawthorn and Pierre Kern. Astrophotonics: a new era for astronomical instruments. *Optics Express*, 17(3):1880, jan 2009. doi: 10.1364/oe.17.001880. URL https://doi.org/10.13642Foe.17.001880.

[36] S. G. Leon-Saval, T. A. Birks, J. Bland-Hawthorn, and M. Englund. Multimode fiber devices with single-mode performance. *Optics Letters*, 30(19):2545–2547, Oct 2005. doi: 10.1364/OL.30.002545.

[37] S. G. Leon-Saval, T. A. Birks, J. Bland-Hawthorn, and M. Englund. Single-mode performance in multimode fibre devices. In *Optical Fiber Communication Conference and Exposition and The National Fiber Optic Engineers Conference*, page PDP25. Optica Publishing Group, 2005. URL https://opg.optica.org/abstract.cfm?URI=OFC-2005-PDP25.

[38] Sergio G. Leon-Saval, Alexander Argyros, and Joss Bland -Hawthorn. Photonic lanterns. *Nanophotonics*, 2:429–440, Dec 2013. doi: 10.1515/nanoph-2013-0035.

[39] Y Sun, T Szkopek, and P.W.E Smith. Demonstration of narrowband high-reflectivity bragg gratings in a novel multimode fiber. *Optics Communications*, 223(1):91–95, 2003. ISSN 0030-4018. doi: https://doi.org/10.1016/

S0030-4018(03)01636-5. URL https://www.sciencedirect.com/science/article/pii/S0030401803016365.

[40] T. A. Birks, B. J. Mangan, A. Díez, J. L. Cruz, and D. F. Murphy. "Photonic lantern" spectral filters in multi-core Fiber. *Optics Express*, 20(13):13996, Jun 2012. doi: 10.1364/OE.20.013996.

[41] Mark Corrigan, Robert J. Harris, Robert R. Thomson, David G. MacLachlan, Jeremy Allington-Smith, Richard Myers, and Tim Morris. Wavefront sensing using a photonic lantern. *SPIE*, 9909:990969, 2016. doi: 10.1117/12.2230568.

[42] Noah R. Van Zandt and Mark F. Spencer. Improved adaptive-optics performance using polychromatic speckle mitigation. *Appl. Opt.*, 59(4):1071–1081, Feb 2020. doi: 10.1364/AO.379972. URL https://opg.optica.org/ao/abstract.cfm?URI=ao-59-4-1071.

[43] T. A. Birks, I. Gris-Sánchez, S. Yerolatsitis, S. G. Leon-Saval, and R. R. Thomson. The photonic lantern. *Advances in Optics and Photonics*, 7(2):107, Jun 2015. doi: 10.1364/AOP.7.000107.

[44] B. S. Kawasaki, K. O. Hill, and R. G. Lamont. Biconical-taper single-mode fiber coupler. *Opt. Lett.*, 6(7):327–328, Jul 1981. doi: 10.1364/OL.6.000327. URL https://opg.optica.org/ol/abstract.cfm?URI=ol-6-7-327.

[45] T.A. Birks and Y.W. Li. The shape of fiber tapers. *Journal of Lightwave Technology*, 10(4):432–438, 1992. doi: 10.1109/50.134196.

[46] J. Bland-Hawthorn, Seong-Sik Min, Emma Lindley, Sergio Leon-Saval, Simon Ellis, Jon Lawrence, Nicolas Beyrand, Martin Roth, Hans-Gerd Löhmannsröben, and Sylvain Veilleux. Multicore fibre technology: the road to multimode photonics. In Ramón Navarro and James H. Burge, editors, *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II*, volume 9912, page 99121O. International Society for Optics and Photonics, SPIE, 2016. doi: 10.1117/12.2231924. URL https://doi.org/10.1117/12.2231924.

[47] Izabela Spaleniak, Nemanja Jovanovic, Simon Gross, Michael Ireland, Jon Lawrence, and Michael Withford. Enabling photonic technologies for seeing-limited telescopes: fabrication of integrated photonic lanterns on a chip. In Ramón Navarro, Colin R. Cunningham, and Eric Prieto, editors, *Modern Technologies in Space- and Ground-based Telescopes and Instrumentation II*, volume 8450, page 845015. International Society for Optics and Photonics, SPIE, 2012. doi: 10.1117/12.925264. URL https://doi.org/10.1117/12.925264.

[48] Sergio G. Leon-Saval, Alexander Argyros, and Joss Bland-Hawthorn. Photonic lanterns: a study of light propagation in multimode to single-mode converters. *Opt. Express*, 18(8):8430–8439, Apr 2010. doi: 10.1364/OE.18.008430. URL http://www.osapublishing.org/oe/abstract.cfm?URI=oe-18-8-8430.

[49] Joss Bland-Hawthorn, S.C. Ellis, Sergio Leon-Saval, Roger Haynes, Martin Roth, H.-G Löhmannsröben, Anthony Horton, Jean-Gabriel Cuby, T.A. Birks, J.S. Lawrence, P Gillingham, Shannon Ryder, and Christopher Trinh. A complex multi-notch astronomical filter to suppress the bright infrared sky. *Nature communications*, 2:581, 12 2011. doi: 10.1038/ncomms1584.

[50] Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42):230–265, 1936. URL http://www.cs.helsinki.fi/u/gionis/cc05/OnComputableNumbers.pdf.

[51] J. von Neumann. *Theory of Self-Reproducing Automata*. University of Illionois Press, Champain, IL, 1966.

[52] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2014.09.003. URL https://www.sciencedirect.com/science/article/pii/S0893608014002135.

[53] Nikolaus Kriegeskorte and Tal Golan. Neural network models and deep learning. *Current Biology*, 29(7):R231–R236, 2019. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2019.02.034. URL https://www.sciencedirect.com/science/article/pii/S0960982219302040.

[54] Hubert L. Dreyfus. *What Computers Still Can't Do*. MIT Press, revised edition edition, 1992. ISBN 978-0-262-04134-8.

[55] Raj Reddy. Foundations and grand challenges of artificial intelligence: AAAI presidential address. *AI Mag.*, 9(4):9–21, 1988. URL https://ojs.aaai.org/index.php/aimagazine/article/view/950.

[56] Warren Mcculloch and Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:127–147, 1943.

[57] Li Yang and Abdallah Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316, 2020. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2020.07.061. URL https://www.sciencedirect.com/science/article/pii/S0925231220311693.

[58] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958. ISSN 0033-295X. doi: 10.1037/h0042519. URL http://dx.doi.org/10.1037/h0042519.

[59] Paul Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Science. Thesis (Ph. D.). Appl. Math. Harvard University*. PhD thesis, School of Mathematic, Harvard University, 01 1974.

[60] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning Representations by Back-propagating Errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0. URL http://www.nature.com/articles/323533a0.

[61] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. doi: 10.1162/neco.1989.1.4.541.

[62] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[63] Xiao-Hu Yu, Guo-An Chen, and Shi-Xin Cheng. Dynamic learning rate optimization of the backpropagation algorithm. *IEEE Transactions on Neural Networks*, 6(3):669–677, 1995. doi: 10.1109/72.377972.

[64] Sebastien Vievard, Steven Bos, Frédéric Cassaing, Alban Ceau, Olivier Guyon, Nemanja Jovanovic, Christoph Keller, Julien Lozi, Frantz Martinache, Aurélie Montmerle, Laurent Mugnier, Mamadou NDiaye, Barnaby Norris, Ananya Sahoo, Jean-Francois Sauvage, Frans Snik, Michael Wilby, and Alisson Wong. Overview of focal plane wavefront sensors to correct for the low wind effect on subaru/scexao. *arXiv: Instrumentation and Methods for Astrophysics*, 12 2019.

[65] Alison P. Wong, Barnaby R. M. Norris, Vincent Deo, Olivier Guyon, Peter G. Tuthill, Julien Lozi, Sébastien Vievard, and Kyohoon Ahn. Machine learning for wavefront sensing. In *Astronomical Telescopes + Instrumentation*, 2022. URL https://api.semanticscholar.org/CorpusID:251962668.

[66] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014. URL http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf.

[67] Paul Fricker. Zernike polynomials, 2021. URL Zernikepolynomials(https://www.mathworks.com/matlabcentral/fileexchange/7687-zernike-polynomials).

[68] von F. Zernike. Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica*, 1(7):689–704, May 1934. doi: 10.1016/S0031-8914(34)80259-5.

[69] Frantz Martinache, Nemanja Jovanovic, and Olivier Guyon. Closed-loop focal plane wavefront control with the scexao instrument. *Astronomy and Astrophysics*, 593:1–8, September 2016. ISSN 0004-6361. doi: 10.1051/0004-6361/201628496.

[70] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike

Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL http://tensorflow.org/. Software available from tensorflow.org.

[71] François Chollet et al. Keras. https://keras.io, 2015.

[72] European Southern Observatory. The vlt's new deformable secondary mirror, 2016. URL https://www.eso.org/public/images/ann16078d/. [Online; accessed December 10, 2022].

[73] E. H. Por, S. Y. Haffert, V. M. Radhakrishnan, D. S. Doelman, M. Van Kooten, and S. P. Bos. High Contrast Imaging for Python (HCIPy): an open-source adaptive optics and coronagraph simulator. In *Adaptive Optics Systems VI*, volume 10703 of *Proc. SPIE*, 2018. doi: 10.1117/12.2314407. URL https://doi.org/10.1117/12.2314407.

[74] Xing Fan. OpticsPy, 2019. URL https://github.com/Sterncat/opticspy.